Introdução à Econofísica

Aula 10

Escalas em dados financeiros

Hoje em dia há uma quantidade imensa de dados financeiros sendo armazenados, negócio a negócio, pelo mundo afora. Gratuitamente, é possível conseguir facilmente dados financeiros diários. Para a modelagem da dinâmica estocástica de preços de um ativo financeiro, podemos utilizar várias escalas de tempo e medir várias funções de preços. Neste curso, utilizamos sempre os preços diários de ações da BM&FBovespa, que também podem ser obtidos gratuitamente. Para a função de preços, de acordo com o que discutimos na quinta aula, adotamos o retorno relativo,

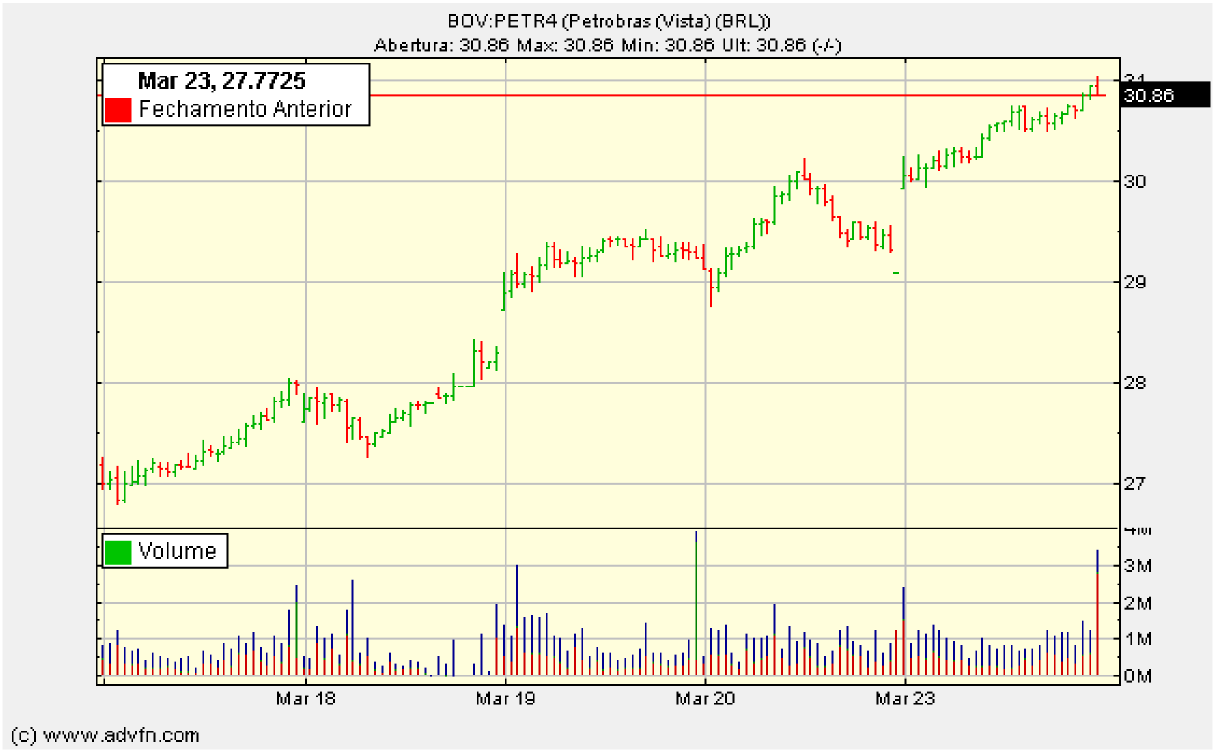
$$R_{i+1} \;=\; \frac{S_{i+1} - S_i}{S_i},$$

e, dependendo da aplicação, o logaritmo do quociente entre o preço de fechamento atual e o anterior,

$$X_{i+1} \;=\; \ln\left(\frac{S_{i+1}}{S_i}\right).$$

Além dessas nossas escolhas, o quinto capítulo do livro-texto cita várias outras que podem ser usadas.

A figura abaixo mostra um gráfico obtido do site muito útil, br.advfn.com. Nesse gráfico, vemos as barras de preços correspondentes a cada quinze minutos de negociação das ações preferenciais da Petrobras. Na parte de baixo da figura, há os volumes de negociação e podemos notar como o final do pregão tipicamente concentra um volume maior do que os outros intervalos de quinze minutos do dia. Assim, a quantidade de negócios varia durante o dia e a distribuição de negócios não é a mesma dia após dia, não sendo, portanto, estacionária.



Processos estocásticos estacionários

A distribuição de probabilidade de preços, em princípio, para ser consistente com essas observações da evolução dos preços, pode ser representada por uma função dependente do preço e do tempo, $f(x,t)$. O valor esperado do preço é, portanto, dado por:

$$E[x(t)] = \int_{-\infty}^{+\infty} dx\, x f(x,t) \, .$$

Seja a função $f(x_1, x_2; t_1, t_2)$ a distribuição de probabilidade conjunta de observar o preço x_1 no instante t_1 e o preço x_2 no instante t_2. A função de autocorrelação é, então, definida como:

$$E[x_1(t_1)\, x_2(t_2)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2\, x_1 x_2 f(x_1, x_2; t_1, t_2) \, .$$

Um processo estocástico geral, para ser completamente caracterizado, requer a função $f(x_1, x_2, \ldots x_n; t_1, t_2, \ldots t_n)$, para todo (x_i, t_i) e n. A maioria dos estudos aborda apenas até a função de dois pontos, $f(x_1, x_2; t_1, t_2)$.

Um processo estocástico $x(t)$ é dito estritamente estacionário se sua distribuição de probabilidades é invariante por translações temporais. Há, porém, outras definições menos restritivas de estacionariedade:

- Definição ampla de processo estacionário:

$$E\left[x\left(t\right)\right] = \mu,$$

$$E\left[x_1\left(t_1\right)x_2\left(t_2\right)\right] = R\left(\tau\right),$$

onde

$$\tau = t_2 - t_1$$

e

$$E\left[x^2\left(t\right)\right] = R\left(0\right).$$

Para essa definição, a variância é independente do tempo.

- Processos estocásticos assintoticamente estacionários ocorrem quando as previsões estatísticas para

$$x_1\left(t_1+c\right), x_2\left(t_2+c\right), \ldots x_n\left(t_n+c\right)$$

não dependem de c se

$$c \rightarrow \infty.$$

- Processos estocásticos estacionários de N-ésima ordem ocorrem quando

$$f\left(x_1, x_2, \ldots x_n; t_1, t_2, \ldots t_n\right) = f\left(x_1, x_2, \ldots x_n; t_1+c, t_2+c, \ldots t_n+c\right)$$

vale apenas para

$$n \leqslant N.$$

- Processos estocásticos estacionários em um intervalo acontecem quando

$$f\left(x_1, x_2, \ldots x_n; t_1, t_2, \ldots t_n\right) = f\left(x_1, x_2, \ldots x_n; t_1+c, t_2+c, \ldots t_n+c\right)$$

para cada t_i e $t_i + c$ dentro do intervalo considerado.

Correlação

A autocorrelação também é denotada por:

$$R\left(t_1, t_2\right) = E\left[x_1\left(t_1\right)x_2\left(t_2\right)\right].$$

Também temos a definição da autocovariância:

$$C\left(t_1, t_2\right) = R\left(t_1, t_2\right) - E\left[x_1\left(t_1\right)\right]E\left[x_2\left(t_2\right)\right].$$

Para processos estacionários, segue:

$$\begin{aligned} C(t_1, t_2) &= C(\tau) \\ &= R(\tau) - \mu^2, \end{aligned}$$

com

$$\tau = t_2 - t_1.$$

Notemos que

$$\begin{aligned} C(0) &= R(0) - \mu^2 \\ &= E\left[x^2(t)\right] - \mu^2 \\ &= \sigma^2, \end{aligned}$$

onde σ^2 é a variância do processo. Para intervalos de tempo muito longos, a autocovariância tende a zero,

$$C(\tau) \approx 0$$

para

$$\tau \rightarrow \infty.$$

Isso acontece porque $x_1(t_1)$ e $x_2(t_2)$ tendem a se descorrelacionar para tempos muito distintos e, portanto, nesse caso:

$$E[x_1(t_1)\, x_2(t_2)] \approx E[x_1(t_1)]\, E[x_2(t_2)].$$

Um típico exemplo é quando o processo tem média nula e

$$R(\tau) = \exp\left(-\frac{\tau}{\tau_c}\right).$$

Nesse caso,

$$\int_0^\infty d\tau \, \exp\left(-\frac{\tau}{\tau_c}\right) = \tau_c$$

e τ_c é chamado de tempo de correlação do processo estocástico, que mede a memória típica do valor da variável x.