

Nome do Aluno: Eduardo Ferrari Magalhães

Nome do Orientador: Raul Ikeda Gomes da Silva

Título: Análise do comportamento de ações utilizando de dados de alta frequência (tick-by-tick)

Palavras Chave: *High Frequency Trading*, *Dados Tick-by-Tick*, *Backtesting*

## 1. Descrição do Problema e Revisão de Literatura

Prever o valor de ações na bolsa de valores é um desafio tão antigo quanto o próprio mercado de ações, essa capacidade, ainda que pouco precisa, permitiria investimentos com muito menos riscos e maior garantia de retorno, fazendo com que o desenvolvimento de recursos nessa área seja o objetivo de diversos pesquisadores.

Apesar de não existirem modelos concretos para prever o mercado (Aldritch, I. 2013), com os avanços computacionais e aumento de dados coletados no dia-a-dia é possível criar novas análises para se obter algumas características dos instrumentos financeiros. A disponibilidade de dados de ultra alta frequência (*Ultra High Frequency Data* ou U-HFD) e do controle do valor de ações em cada evento (*tick-by-tick*) são os fatores chave para uma nova abordagem, e para completar esse desafio.

A captação de dados em ultra alta frequência é definida pela primeira vez por Engle, R. (1996) como a captação de dados de mercado de forma que todas as transações sejam gravadas ainda que intervalos de tempo irregulares. Dessa forma o valor das ações não é controlado em uma escala de tempo e sim em relação ao número de transações que aconteceram naquele período de funcionamento da bolsa, esse número é denominado *tick*.

Com a utilização desses dados, é possível formar análises que observam características comumente descartadas, alimentando uma nova classe de modelos que utilizam dezenas de milhares de dados de transações por dia. Esse fator é relevante, pois dados de duração muito longa não transmitem informações úteis. (Easley D. & O'Hara M., 1992; Engle, R & Patton A. 2004; Schotman P., Frijns B. 2004).

Ainda assim, o maior desafio para a modelagem de uma determinada ação não é efetivamente encontrar um modelo que se adapte aos dados coletados e sim garantir que esse modelo consegue se adaptar a dados de fora da amostra escolhida (*Out Of Sample*) como avaliado para diversos modelos já existentes em (Rogers L. & Zane O. 1998).

Nesse contexto é importante entender que modelar algumas ações de um mercado específico e esperar que ele se aplique a todos os mercados é irreal (Engle, R., Sun Z., 2005). Os diferentes mecanismos de investimento e composição do mercado, criam relações de variação de preço e volume de investimentos muito diferentes para os mercados.

## 2. Objetivo

O presente projeto apresentará uma análise de dados de alta frequência, restritos às ações da Bolsa de Valores de São Paulo. A análise visa testar aderência de diferentes modelos clássicos (Regressão Linear, Modelos Dinâmicos Discretos, etc.) assim como alguns modelos de *Machine Learning* (SVM, *Random Forests*, etc.) para tentar descrever o comportamento dos dados.

### 3. Metodologia (Proposta)

Com o objetivo de modelar os dados de Ultra Alta Frequência, amostras serão coletadas a partir de provedores de dados da Bloomberg de forma que uma base de dados seja construída e uma série de modelos estatísticos possam ser testados nesse grupo de dados.

Dada a natureza discreta dos dados, será necessário aplicar alguns filtros para atenuação de ruídos, antes de se aplicar as técnicas de modelagem. Em seguida, a previsibilidade dos modelos será validada através de métricas para trading de alta frequência de forma a verificar sua aderência.

Para tornar o modelo mais flexível, diferentes ações selecionadas por sua representatividade no mercado (Índices ponderados de capitalização de mercado). De forma semelhante ao artigo de Palágyi Z & Mantegna R. (1998) que utiliza as quatro principais ações do mercado de Budapeste para criar um modelo mais confiável de maturidade do mercado ao longo do tempo. A partir disso é possível mapear, por exemplo, o volume de transações de cada ação e através disso estimar os preços das mesmas no decorrer do período. (Amihud Y., 2002; Lesmond D., 2005; Liu W., 2006).

Os dados serão divididos em treinamento e testes de forma a não permitir o *overfitting*, assim é possível evitar que o modelo sofra o efeito descrito em Rogers L. & Zane O. (1998) de não funcionar fora da amostra inicial.

Essas amostras serão compostas das principais ações do índice Bovespa de forma a obter o maior número de dados para alimentar o sistema. Por exemplo as ações com maior média de negociações no período de 07/05 até 18/05 estão apresentadas na Tabela 1.

Tabela 1: Relevância de diversas ações na Bovespa

Ação	Número de Negócios	% do total
Total	1.505.490,00	100,00%
PETR4	81.849,20	5,44%
VALE3	36.411,20	2,42%
ABEV3	32.354,80	2,15%
BBDC4	31.403,70	2,09%
ITUB4	30.137,80	2,00%
BBAS3	25.170,70	1,67%

Dessa forma, ações listadas contemplam mais de 15% de todas as negociações do mercado no período analisado.

### 4. Resultados Esperados

Através das medidas de *backtesting*, serão produzidos resultados comparativos entre os modelos usando dados de alta frequência com diferentes períodos e diferentes calibrações verificando se o fenômeno apontado na literatura se mantém para o mercado brasileiro. Todavia, como já discutido na literatura, não existem garantias de previsibilidade dos modelos, sobretudo para longos períodos de tempo.

## 5. Referências Bibliográficas

Aldritch, I., “High Frequency Trading”, 2013. 2a Edição, Wiley Trading Series.

Engle, R., “The Econometrics Of Ultra High Frequency data”, 1996, National Bureau of Economic Research.

Easley, D., and M. O’Hara, “Time and the Process of Security Price Adjustment”, 1992, Journal of Finance, Volume 47, Páginas 905–927.

Engle, R., and A. J. Patton, “Impacts of Trades in an Error-Correction Model of quote prices,”, 2004, Journal of Financial Markets, Volume 7, Páginas 1–25.

Schotman P., Frijns B. “Price Discovery in Tick Time”, 2004, Journal of Empirical Finance Volume 16-5, December 2009, Páginas 759-776

Rogers, L., Zane, O. “Designing and estimating models of high-frequency data”, 1998, University of Bath preprint

Engle, R., Sun Z., “Forecasting Volatility Using Tick by Tick Data”, 2005, SSRN Electronic Journal

Palágyi Z., Mantegna R., “Empirical investigation of stock price dynamics in an emerging market”, 1999, Physica A: Statistical Mechanics and its Applications Volume 269-1, Páginas 132-139.

Amihud Y., “Illiquidity and stock returns: cross-section and time-series effects”, 2002, Journal of Financial Markets Volume 5, Páginas 31-56

Lesmond D., “Liquidity of emerging markets”, 2005, Journal of Financial Economics Volume 77, Páginas 411-452

Liu W., “A liquidity-augmented capital asset pricing model”, 2006, Article in Journal of Financial Economics Volume 86, Páginas 631-671

## Cronograma de atividades

Período: 2018-2

Atividades	2018					2019						
	ago	set	out	nov	dez	jan	fev	mar	abr	mai	jun	jul
Estudo do Problema	X	X										
Coleta de Dados		X	X	X								
Análise dos dados				X	X	X	X	X	X			
Análise de Resultados								X	X	X	X	
Redação de Relatório Final											X	X