



**UNIVERSIDADE
FEDERAL DO CEARÁ**
CAMPUS SOBRAL

Curso de Engenharia Elétrica

Processamento Digital de Sinais

Predição de Stock Market

Alunos: Julio Cesar Ferreira Lima - 393849

Profa: Alexandre Fernandes

Sobral, **Dezembro** de **2018**.

1. INTRODUÇÃO E JUSTIFICATIVA

Stock Market é um dos maiores mercados do mundo, com suas negociações de parte de empresas e investimentos por meio de pequeníssimas parcelas, torna capaz qualquer um de desempenhar um papel de compra e venda delas obtendo assim algum lucro com a negociação. O fato de o preço das ações ser bastante acessível e haver um certo frenesi na obtenção de dinheiro a partir de negociações, tomando como base apenas o próprio dinheiro e expertise de análise de gráficos, trouxe ao pública certa notoriedade. Desse modo os cientistas logo desempenharam técnicas baseadas em procedimentos análogos a acontecimentos na natureza e a matemática para tentar prever tais acontecimentos dentro do mercado de ações.

Este trabalho baseia em ferramentas descobertas por Joseph Fourier, onde a mais importante delas, para este trabalho será a Transformada Discreta de Fourier. A Transformada Discreta de Fourier é análoga a contínua e tem a função de para sinais de N amostras igualmente espaçadas por um tempo T.

$$F[n] = \sum_{k=0}^{N-1} f[k] e^{-j \frac{2\pi}{N} nk} \quad (n = 0 : N - 1) \quad \text{Eq. (1)}$$

Desse modo deve ser considerado que por ser obtido apenas N amostras a Transformada Discreta de Fourier será simétrica e periódica.

Outra variação da Transformada de Fourier é a FFT (Fast Fourier Transform), que é muito utilizada em análises financeiras e com base nisso o indivíduo baseará seu trade nesse sinal de predição obtido. Essas estratégias são usadas por muitos grupos de Bancos de Investimentos, como Morgan Stanley, Goldman Sachs.

Nesse relatório também foi utilizada uma técnica utilizada na estatística de modo a obter a curva que mais se aproxima de um padrão de amostragem de dados. A priori tal estratégia pode não parecer uma boa estratégia relacionada à inteligência computacional, no entanto por descrever o comportamento da maioria da população de amostras, demonstra tendência o que pode ser considerado aprendizado. Defendido pelo Teorema Gauss-Markov que garante confiabilidade não enviesada de mínima variância linear e desenvolvido por Carl Friedrich Gauss, quando tinha apenas dezoito anos, aqui, sua qualidade de estimativa será medida pelo coeficiente de determinação, também chamado de R^2 , que é uma medida de ajustamento de um modelo estatístico linear generalizado assim como um variação desse parâmetro ajustado, que considera a qualidade das amostras e não só totalmente sua quantidade, retirando assim a importância do ruído nas amostras. Todo esse trabalho será usada variâncias dessas teorias, variando apenas a dimensionalidade do sistema e a ordem da estimativa.

Ao se aproximar de uma curva, é natural que exista um erro associado, assim temos essa aproximação dada por:

$$y = \alpha + \beta x + \epsilon, \quad \text{Eq. (2)}$$

Sendo “a” e “b” parâmetros de aproximação, “e” o erro associado, “y” a saída da instância de “x”.

Discorrendo sobre a reta poderemos fazer o somatório dos erros dado um mesmo conjunto de parâmetros “a” e “b”. Tais erros podem ser somados e será mínimo quando o \hat{fit} da curva for ótimo. Considerando que há uma curva onde a combinação dos parâmetros “a” e “b” são entradas e que é igual a soma dos erros totais, temos:

$$S(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad \text{Eq. (3)}$$

Tomando a derivada parcial de cada um dos parâmetros e determinando a função para que essa variável seja zero, dado que nesse ponto será a maximização do *fit* e logo minimização do erro associado a aquele parâmetro, temos:

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial x} * \frac{\partial x}{\partial a} \quad \text{Eq. (4)}$$

Após alguns algebrismos para a determinação de “a” e “b”:

$$a = \bar{y} - b\bar{x} \quad \text{Eq. (5)}$$

$$b = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} \quad \text{Eq. (6)}$$

Extrapolando a Regressão Linear Simples para a Regressão Múltipla, temos então a possibilidade de multidimensionalidade e aumento de ordem do aproximador, dado que não há apenas mais um “x” e sim um “xi”, além de que há a soma e atribuições de pesos para variável aleatória “xi” e há progressão em graus, não limitando-se mais apenas ao primeiro.

Extrapolando em relação a dimensionalidade:

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \dots + x_k\beta_k + \varepsilon \quad \text{Eq. (7)}$$

Extrapolando em relação ao número de graus, temos que valor de uma mesma coluna em “xi” correspondem ao mesmo grau e ele crescer da esquerda para a direita.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & \dots & x_{k3} \\ 1 & x_{14} & x_{24} & \dots & x_{k4} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix} \times \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ \dots \\ e_n \end{pmatrix} \quad \text{Eq. (8)}$$

Simplificando tudo em matrizes, voltamos simplicidade da tomada inicialmente:

$$y = Xb + e \quad \text{Eq. (9)}$$

Tomando posse de mais alguns algebrismos para a simplificação da extração de beta, temos:

$$\begin{aligned} S(b) &= (y - Xb)'(y - Xb) \\ &= y'y - y'Xb - b'X'y + b'X'Xb \end{aligned} \quad \text{Eq. (10)}$$

$$\begin{aligned}
\frac{\partial S}{\partial b} &= -2X'y + 2X'Xb = 0 \\
X'Xb &= X'y \\
b &= (X'X)^{-1}X'y
\end{aligned}
\tag{Eq. (11)}$$

Objetivando agora a quantização do *fit* curva, faz-se uso do do coeficiente de determinação:

$$SQ_{\text{tot}} = \sum_{i=1}^n (y_i - \bar{y})^2 \tag{Eq. (12)}$$

Onde esta última equação é dada como o somatório da diferença da distância média entre a curva e a média desta ao quadrado visando apenas tomar posse de uma ordem desse valor de forma absoluta.

Assim como, faz-se necessário a tomada do somatório das diferenças entre a curva e o *fit*:

$$SQ_{\text{res}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{Eq. (13)}$$

Por fim R^2 pode ser determinado por:

$$R^2 = \frac{SQ_{\text{exp}}}{SQ_{\text{tot}}} = 1 - \frac{SQ_{\text{res}}}{SQ_{\text{tot}}} \tag{Eq. (14)}$$

Por último, visando combater a negação do princípio da parcimônia por R^2 , que dá notoriedade a dados de com baixo índice de pertencimento a curva, aplica-se o método do R^2 ajustado:

$$\bar{R}^2 = 1 - \frac{n-1}{n-(k+1)} (1 - R^2) \tag{Eq. (15)}$$

Onde “n” é a quantidade de dados e “k” a ordem adotada.

2. OBJETIVOS

3.1 GERAIS

Validar o algoritmo desenvolvido pelo autor ao qual este trabalho se baseia [1], de forma a apresentar uma nova possibilidade de predição de mercado utilizando o algoritmo da transformada rápida de fourier.

3.2 ESPECÍFICOS

Tratar diferentes sinais de mercado a fim de verificar a resposta produzida pela predição utilizando este algoritmo, descobrindo então situações de uso, possíveis falhas e utilidades em geral para a ferramenta.

3. METODOLOGIA

O método que este trabalho se baseia, utiliza como base a regressão com mínimos quadrados uma linha média que melhor se adequa aos dados, especificamente podendo chegar até a quinta ordem (a utilização em quinta ordem já fora suficiente para os dados analisados). Idealmente a ordem utilizada poderia ser tão grande quanto fosse necessário, mas visando a generalização do algoritmo

fora travado em quinta ordem.

Como fora desejada a experimentação e teste da acurácia do algoritmo não foram inseridos no algoritmo todas as amostras, deixando assim as amostras do final destinadas a serem preditas e assim comparado.

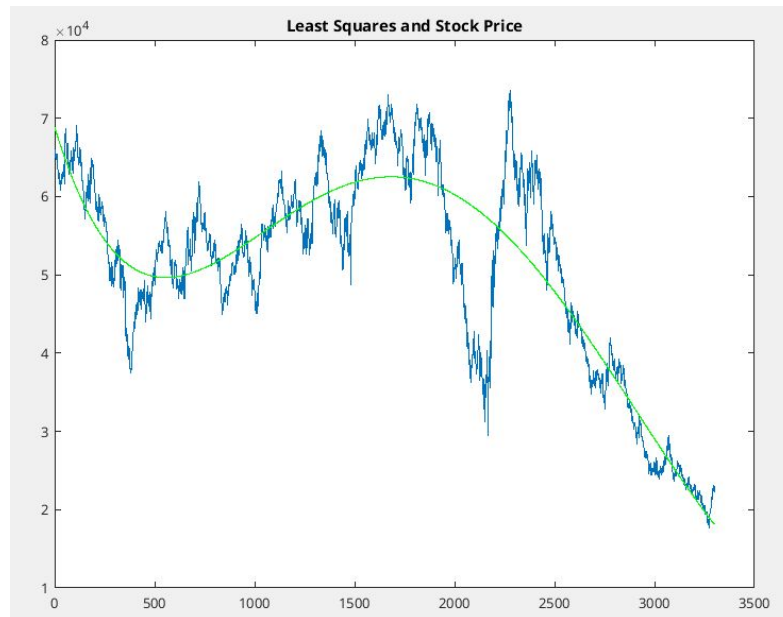


Figura 1 – Gráfico da ação IBOV e o polinômio de 5 ordem.

O próximo ponto do algoritmo fora extrair a diferença entre a curva IBOV e a regressão.

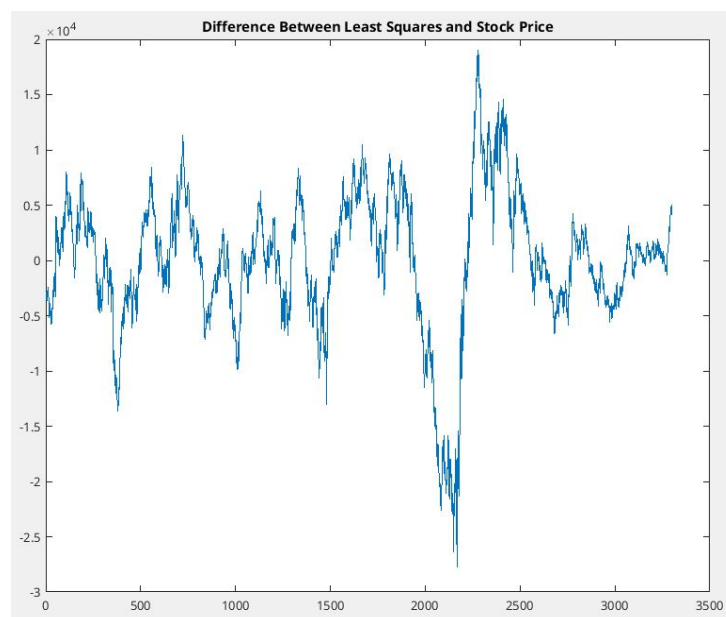


Figura 2 – Gráfico diferença entre IBOV e o polinômio de 5 ordem.

Após esse passo de fato irá utilizar-se as ferramentas de Fourier, por meio da aplicação do módulo da FFT no sinal obtido anteriormente. Isto é, a representação de uma série temporal em frequência de curvas de senoidais.

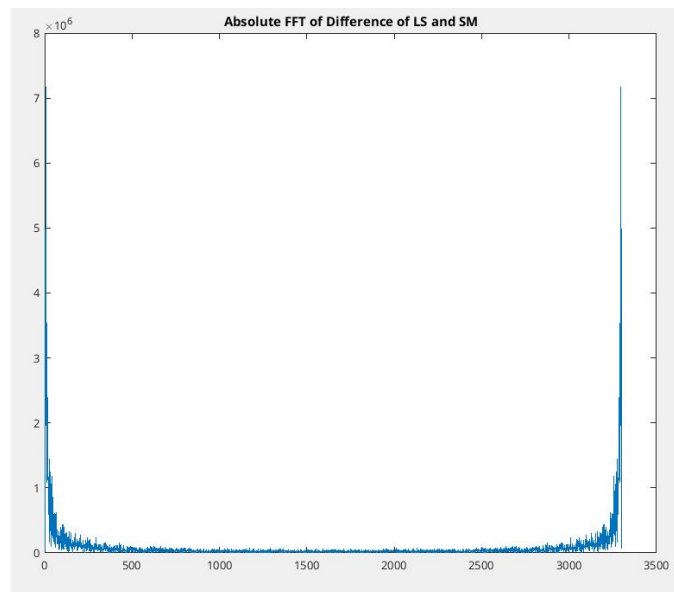


Figura 3 – Gráfico do módulo da FFT da diferença entre IBOV e o polinômio de 5 ordem.

O próximos e último passo da predição será a interpolação dos dados desejados dentro da próxima janela de tempo tomando como base para isso a Transformada Fourier Discreta, da forma como o autor em que esse trabalho se baseia[1]. Onde o sinal $YY(n)$ será somado a regressão.

```

a(k)=real(Y(k));
b(k)=-imag(Y(k));
omk=2*pi*(k-1)/365;
YY(n)=YY(n)+a(k)*cos(omk*(n-1))+b(k)*sin(omk*(n-1));

```

Figura 3 – Algoritmo para obtenção da Transformada Fourier Discreta.

Com o objetivo de comparar a estratégia com a resposta real do sistema, fora implementado o coeficiente de determinação R^2 e R^2 ajustado.

4. RESULTADOS

Analisando as ações da IBOV para predição de 100 amostras fora obtido uma diferença de:

Sem o algoritmo:

- Determination Coefficient: 80.248%
- Determination Coefficient Adjusted: 80.2189%

Com o algoritmo:

- Determination Coefficient: 82.9265%
- Determination Coefficient Adjusted: 82.9021%

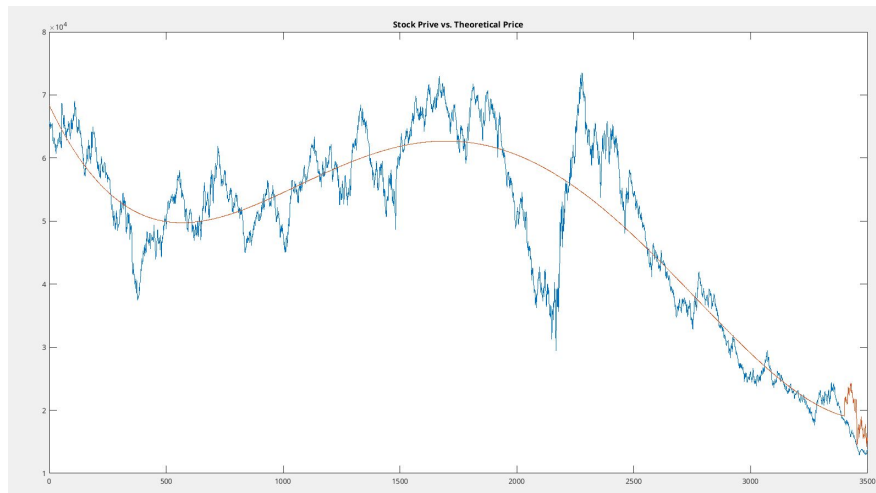


Figura 4 – Predição do índice IBOV com 100 amostras.

Novamente para as ações da IBOV para predição de 350 amostras fora obtido uma diferença de:

Sem o algoritmo:

- Determination Coefficient: 77.3224%
- Determination Coefficient Adjusted: 77.2863%

Com o algoritmo:

- Determination Coefficient: 76.1609%
- Determination Coefficient Adjusted: 76.1268%

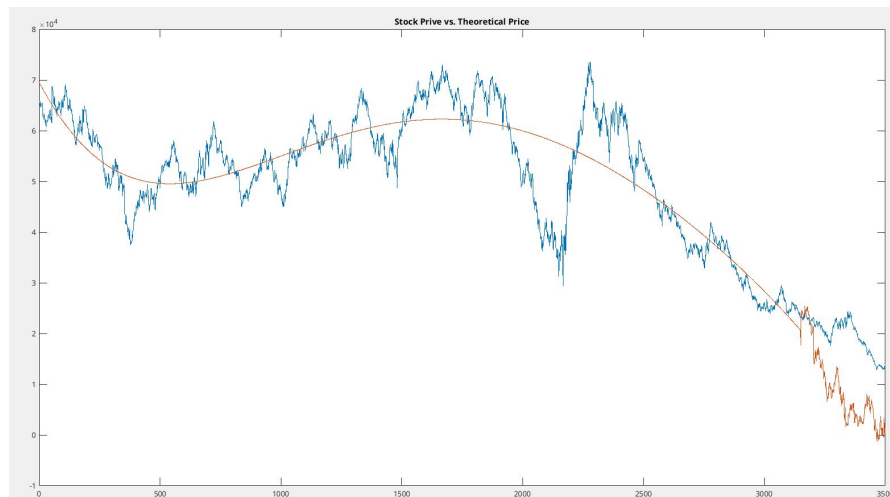


Figura 5 – Predição do índice IBOV com 350 amostras.

Já o índice PETR4 para predição de 100 amostras fora obtido uma diferença de:

Sem o algoritmo:

- Determination Coefficient: 79.5299%
- Determination Coefficient Adjusted: 79.4997%

Com o algoritmo:

- Determination Coefficient: 80.524%
- Determination Coefficient Adjusted: 80.4961%

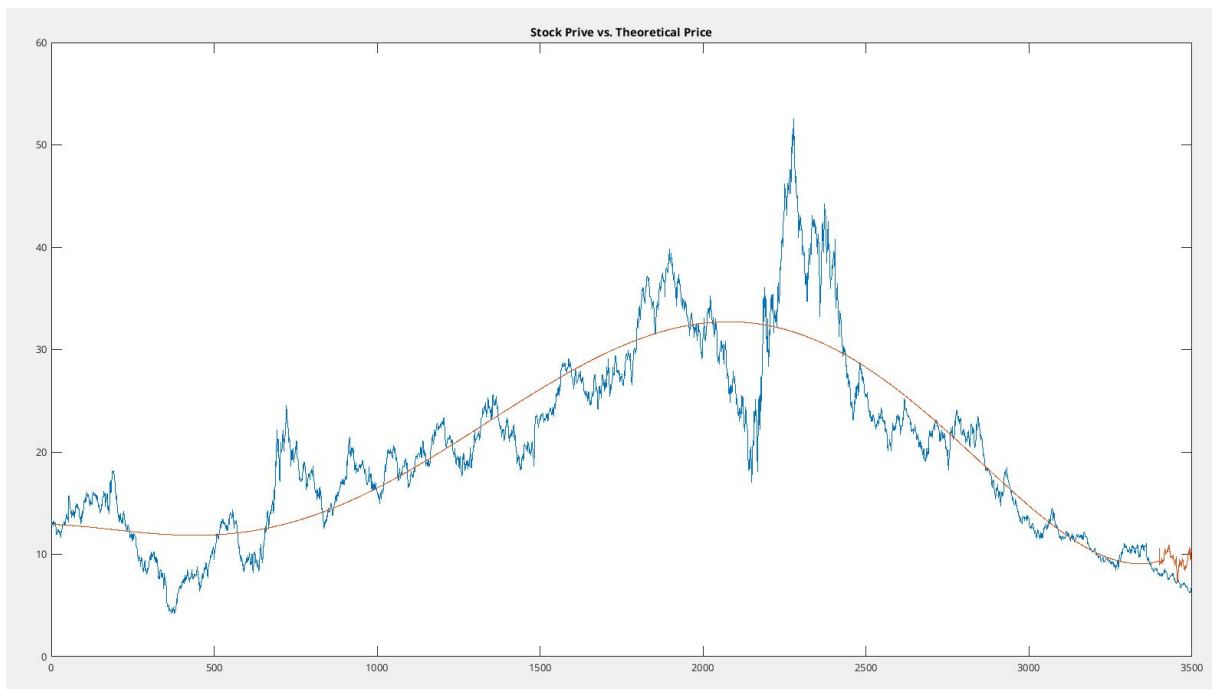


Figura 6 – Predição do índice PETR4 com 100 amostras.

No entanto para alguns gráficos a predição foi terrivelmente inexata, dispensando até a análise de por fator de ajuste.

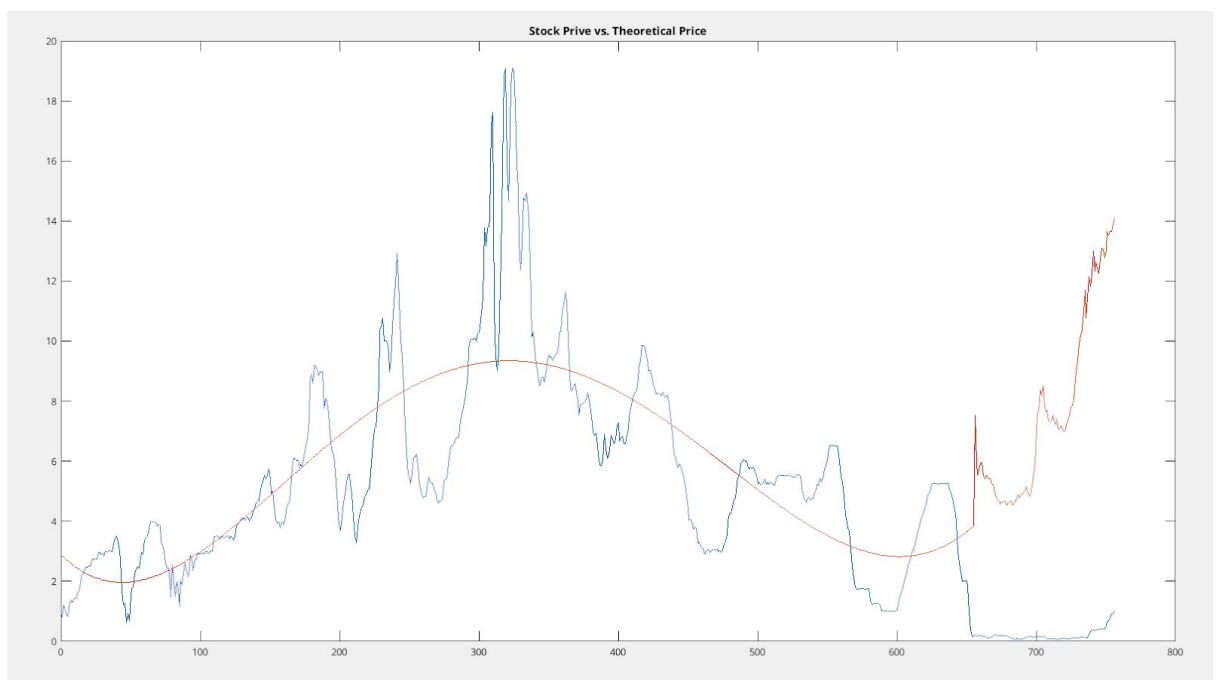


Figura 7 – Predição do índice FEDFUNDS com 100 amostras.

6. CONCLUSÃO

Este trabalho mostrou-se bastante satisfatório na predição de alguns sinais de mercado, como fora o caso das ações PETR4 e IBOV. No entanto em alguns sinais seu uso deverá ser guiado pelo usuário de modo a melhorar sua acurácia.

Possíveis adequações para o software desenvolvido neste trabalho seria adicionar a funcionalidade de mudar a janela de amostras, utilizando assim a predição por esse modelo em sinais que tivesse determinado nível de tendência, percentualmente em relação ao número de amostras preditas. Isso por conta que em um longo prazo uma quantidade de amostras de frequências já não deveriam influenciar na predição, no entanto sem que haja esse controle influenciarão. Desse modo o algoritmo melhoraria em caso de ponderamento das amostras por algum outro princípio que pudesse prever até que ponto é interessante considerar os dados.

Por fim aconselha-se a utilização desse algoritmo não considerando como uma certeza, dada a complexidade de um sinal de mercado, mas sim como um sinal onde se pode extrair possibilidades e assim correlacionar com outros, como médias móveis e etc, durante o trabalho do trade.

5. BIBLIOGRAFIA

- [1] WILLIAMS, Michael Barrett ; JANI, Soham; PATHAK, Mayank. **Fast Fourier Transform in Predicting Financial Securities Prices**. Disponível em: <<https://www.math.utah.edu/~gustafso/s2016/2270/published-projects-2016/williamsBarrett/williamsBarrett-Fast-Fourier-Transform-Predicting-Financial-Securities-Prices.pdf>>. Acesso em: 20 de Novembro de 2018.
- [3] JUNIOR, Jarbas. **Notas de Aula: Regressão Linear Múltipla** . Acesso em abril de 2018.
- [4] JUNIOR, Jarbas. **Notas de Aula: Regressão Simples** . Acesso em abril de 2018.