UNIVERSIDADE DE SÃO PAULO

RODOLFO LIMA DA SILVA, Nº 8954691

SME0808– SÉRIES TEMPORAIS E APRENDIZADO DINÂMICO

PROF. RICARDO SANDES EHLERS

PROJETO 1 - ANÁLISE DE ALGUMAS SÉRIES TEMPORAIS

SÃO CARLOS

1. Introdução e Objetivo

Uma série temporal refere-se a uma coleção de dados observados sequencialmente ao longo do tempo. Através delas, de suas observações, características e estruturas, é possível tirar valiosas informações do seu comportamento e até fazer predições com base em dados passados.

Para ilustrar conceitos introdutórios de análise de séries temporais, foram utilizados duas bases para extrair algumas coleções de dados temporais. Dados coletados pelo laboratório de monitoramento global NOOA e dados financeiros retirados do Yahoo Finance.

2. Metodologia

Para o propósito do projeto foi usado duas bases de dados já mencionadas, e para a análise e visualização das séries temporais foram utilizados a linguagem open source Python e seus pacotes disponíveis. Sendo mais específico, foi utilizado o pacote Pandas para manipulações e preparações dos dados, utilizou-se Yfinance para fornecer dados financeiros de maneira rápida e simples, Scipy e Statsmodel para testes e análises estatísticas. e por fim, Matplotlib e Seabor para visualização dos dados.

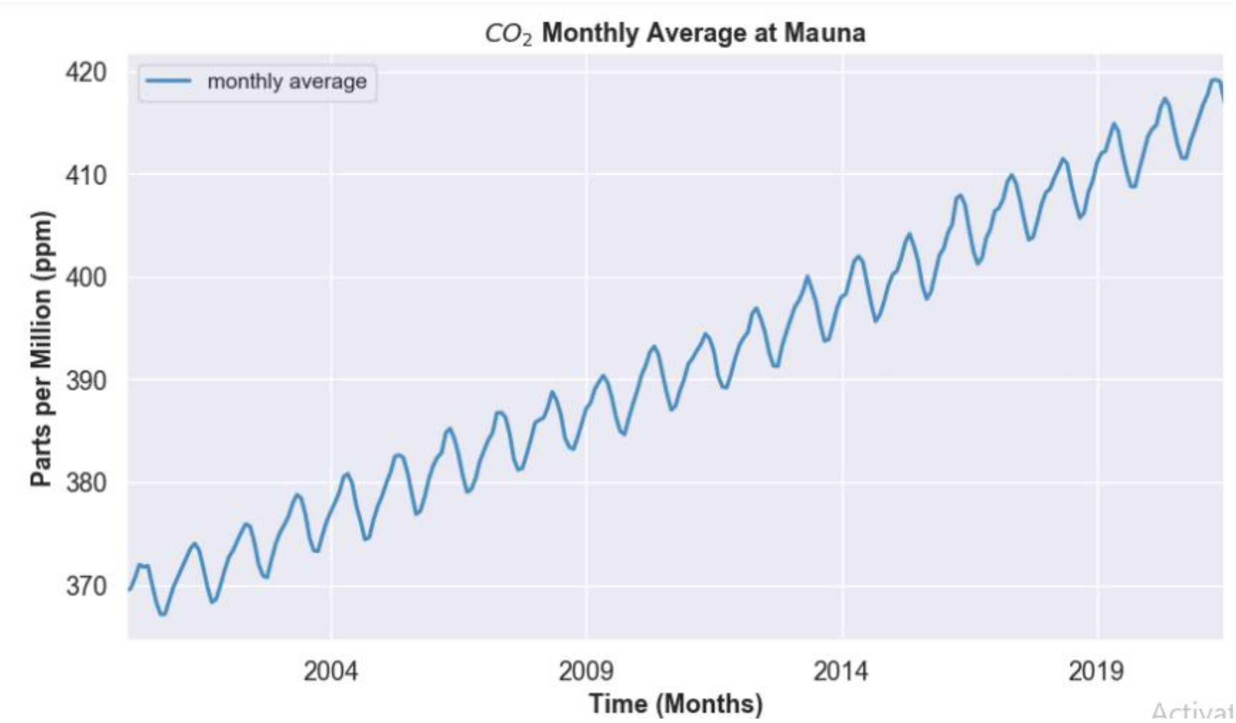
O script do código em Pyhton aplicado ao projeto estará anexado no final desse documento.

3. Discussão dos resultados

3.1 Concentração Atmosférica de CO2 e N2O

A seguinte figura 3.1 refere-se à concentração atmosférica de CO_2 em Mauna Loa Observatory, Hawai USA.

Figura 3.1 - Concentração mensal média de CO_2 em Mauna de 2000 a 2021



Fonte - Elaboração propria com base nos dados da National Oceanic and Atmospheric Administration (NOAA)

A série possui uma estrutura sazonal e de tendência, o que viola a definição de estacionaidade. No entanto, é possível identificar que a série possui dois componentes que podem ser aproximados por curvas já conhecidas. O componente de tendência da série pode ser aproximado por uma curva linear pelo método de regressão linear por mínimos quadrados, generalizando, uma regressão que aproxima a série a um polinômio de grau n:

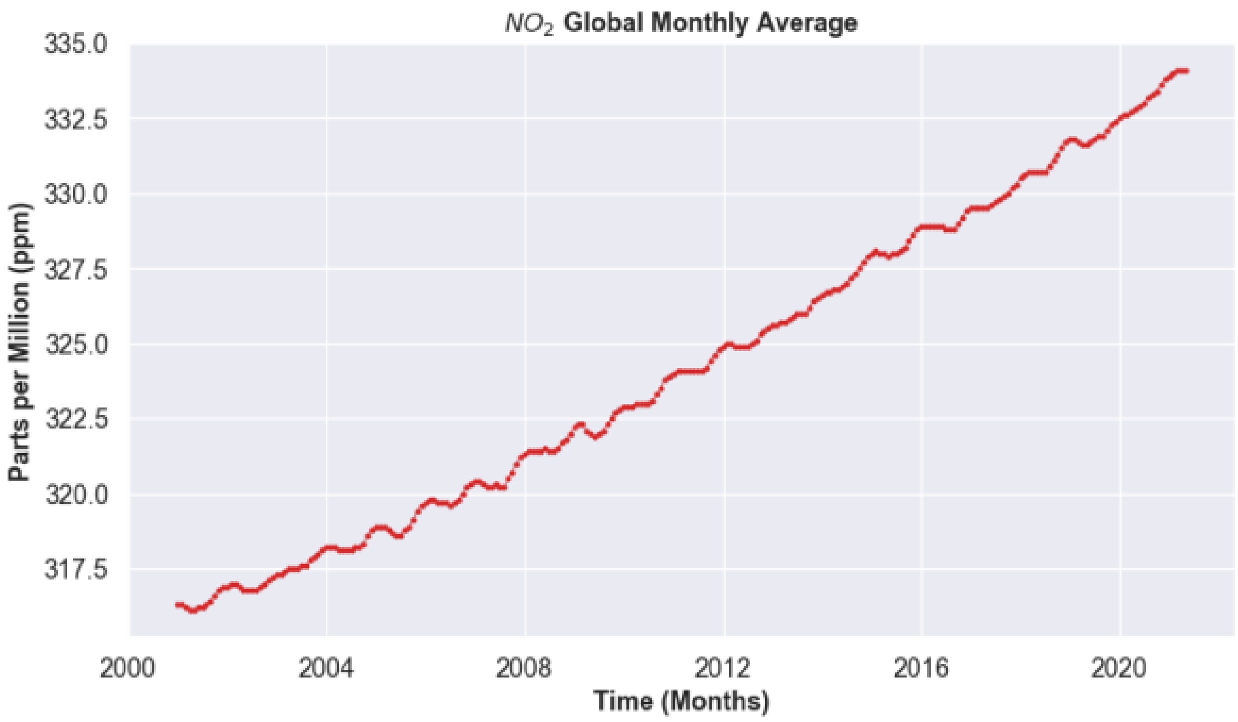
$$y_t = \beta_0 + \beta_1 t + \cdots + \beta_n t^n + \epsilon_t$$

Já o componente sazonal pode ser aproximado por um modelo de regressão harmônico, ou seja, uma equação periódica, pelo mesmo método dos mínimos quadrados.

$$y_t = a_1 sen(2\pi\omega_1 t) + b_1 cos(2\pi\omega_1 t) + \cdots + a_n sen(2\pi\omega_n t) + b_n cos(2\pi\omega_n t) + \epsilon_t$$

O gráfico da figura 3.2 mostra a concentração atmosférica global de NO2.

Figura 3.2 - Média Global de Concentração Atmosférica de NO2 em partes por milhões. Dados de 2000 a 2021



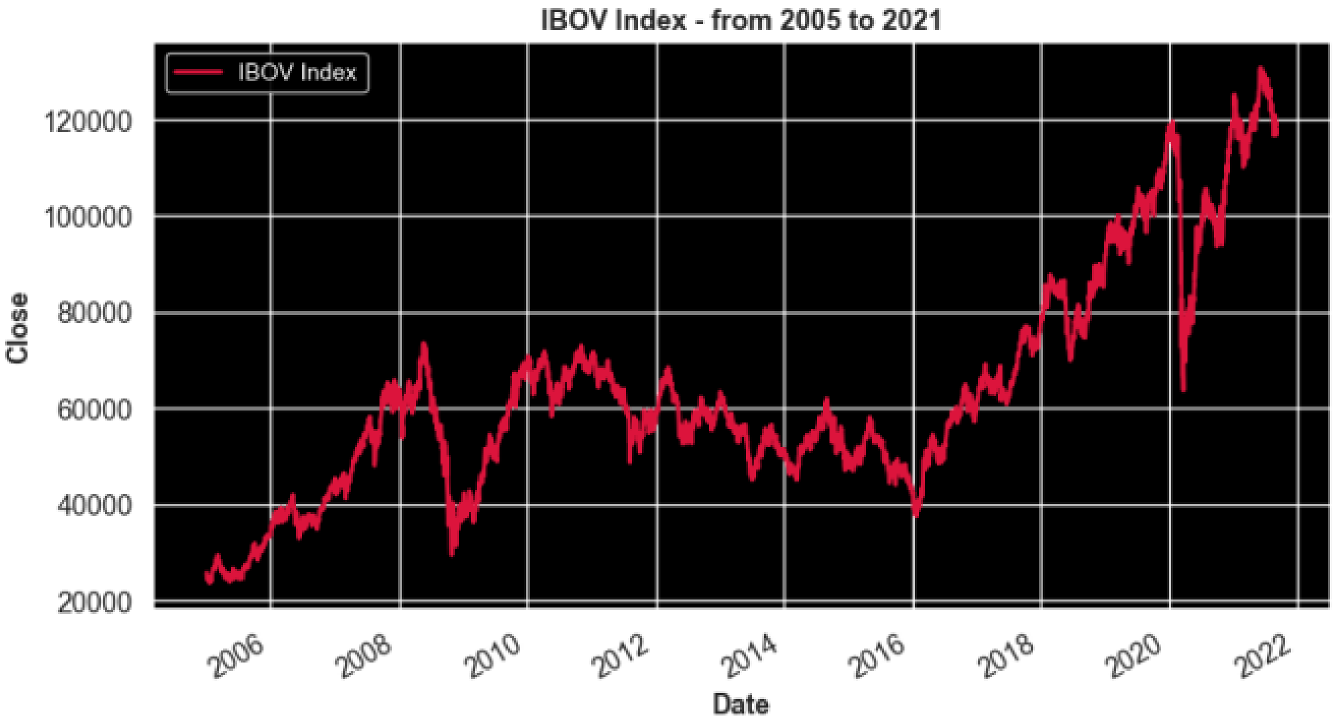
Fonte - Elaboração própria com base nos dados da National Oceanic and Atmospheric Administration (NOAA)

Na figura acima é notório que a série tenha uma estrutura não-estacionária de tendência. Dado a característica da curva, pode-se afirmar que um dos métodos adequados de se modelar a séria é aproximar a um polinômio de grau n pelo método de regressão por mínimos quadrados

3.2 Índice IBOVESPA

Análise de séries temporais é um campo de ímpar relevância para aplicações em finanças, uma vez que os dados da maioria das cotações de ativos financeiros, índices e variáveis econômicas são fornecidos em relação ao tempo. Um indicador financeiro muito importante para o Brasil é o índice IBOVESPA. Sua cotação diária pode ser visualizada pela figura 3.3.

Figura 3.3 - Fechamento diário do índice IBOVESPA

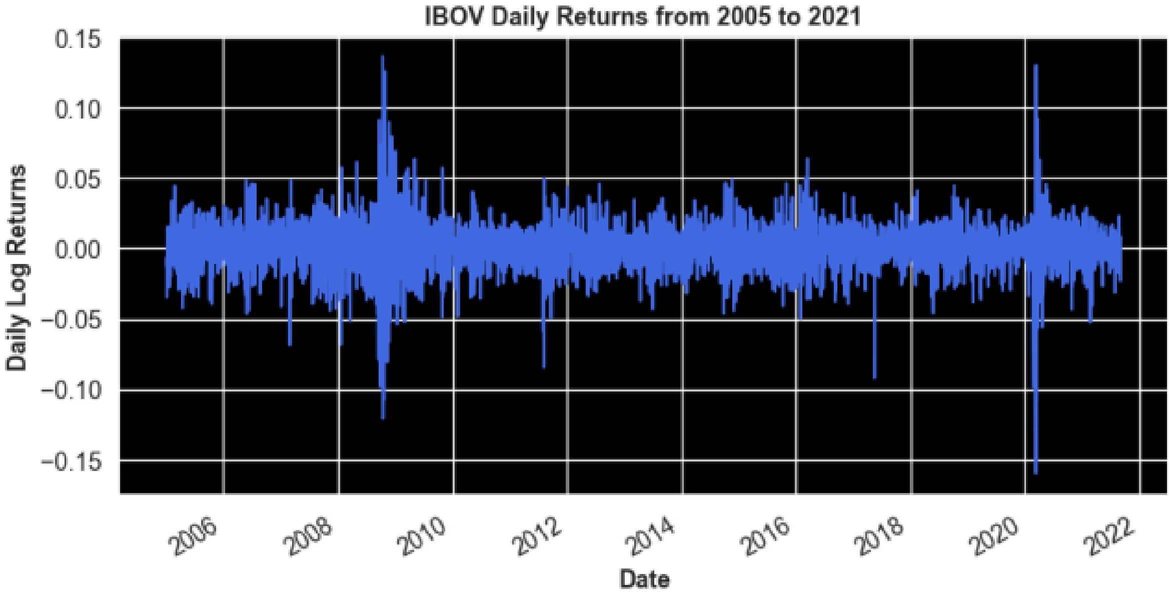


Fonte - Elaboração própria com base nos dados do Yahoo Finance.

O fechamento diário do índice ao longo do tempo, assim como cotações de ações na Bolsa de valores, podem ser representados como um passeio aleatório de média $E(y_t) = \mu$ e variância $Var(y_t) = \sigma^2$. O passeio aleatório é um processo discreto aleatório cuja parâmetros como média, variância e auto variâncias dependem do tempo t, o que infringe o requisito para série estacionária.

Muitas vezes é mais interessante modelar o retorno esperado e a variância do retorno. O gráfico dos retornos do índice IBOVESPA pode ser observado na figura 3.4.

Figura 3.4 - Retornos diários do índice IBOVESPA

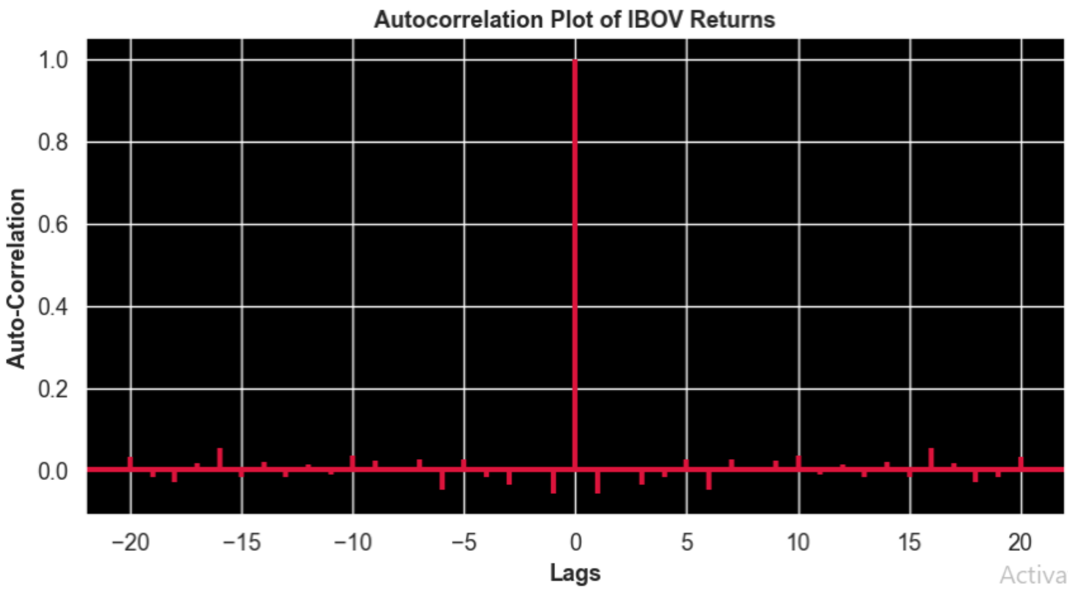


Fonte - Elaboração própria com base nos dados do Yahoo Finance

Observando a figura 3.4 é plausível considerar que se trata de um processo puramente aleatório e estacionário, uma vez que os retornos do índice parecem flutuar aleatoriamente ao redor de 0 e a variância, média parecem serem constante durante todo o tempo registrado, podemos dizer o mesmo para a auto correlação também. Isso significaria que a variável aleatória retorno diário do índice é independente entre seus retornos.

Para verificar com mais acurácia, observemos o gráfico de correlação (correlograma) dos retornos ilustrado pela figura 3.5

Figura 3.5 - Auto correlação dos retornos diários do índice IBOVESPA



Fonte - Elaboração própria com base nos dados do Yahoo Finance

Através do correlograma podemos acatar as suposições iniciais de estacionariedade fraca, afinal, podemos verificar com certa confiança as seguintes propriedades de um processo estocástico estacionário:

$$p(k) = \begin{cases} 1, & \text{if } k = 0 \\ 0, & \text{if } k = \pm 1, \pm 2, \cdots \end{cases}$$

- $-1 < p(k) > +1$
- $p(k) = p(-k)$

Onde p(k) é função de auto-correlação da série. É possível ser mais rigoroso ao avaliar a estacionariedade de um processo estocástico e aplicar o teste de hipótese para estacionariedade, conhecido como teste de Adfuller. O resultado do teste está presente na figura 3.6 logo abaixo:

Figura 3.6 - Output do teste Adfuller aplicado aos retornos do índice IBOVESPA entre 2005 à 2021

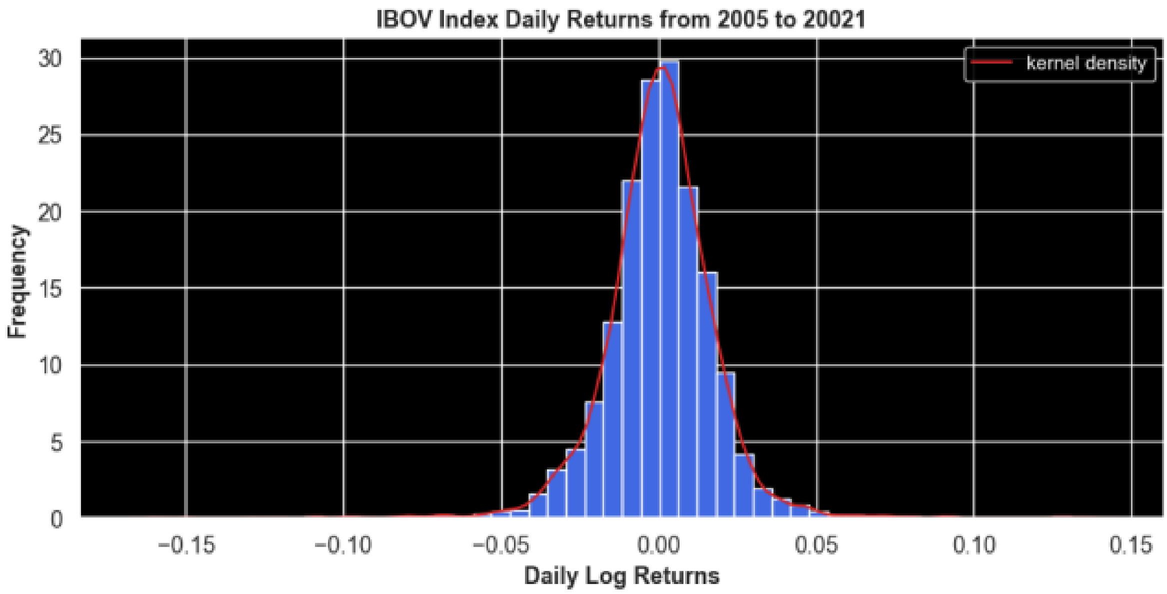
```
ADF Statistic: -13.390277
p-value: 0.000000
Critical Values:
	1%: -3.432
	5%: -2.862
	10%: -2.567
Reject the null hypothesis (H0), the data does not have a unit root and is stationary.
```

Fonte - Elaboração própria, utilizando a linguagem Python para aplicação do teste.

Dessa forma, podemos afirmar com 95% de confiança que a série dos retornos do índice IBOVESPA segue uma estrutura estacionária.

Análises explanatórias e descritivas dos dados também são úteis para descrever propriedades e características da série temporal, como por exemplo a distribuição dos retornos do índice, evidenciado na figura 3.7.

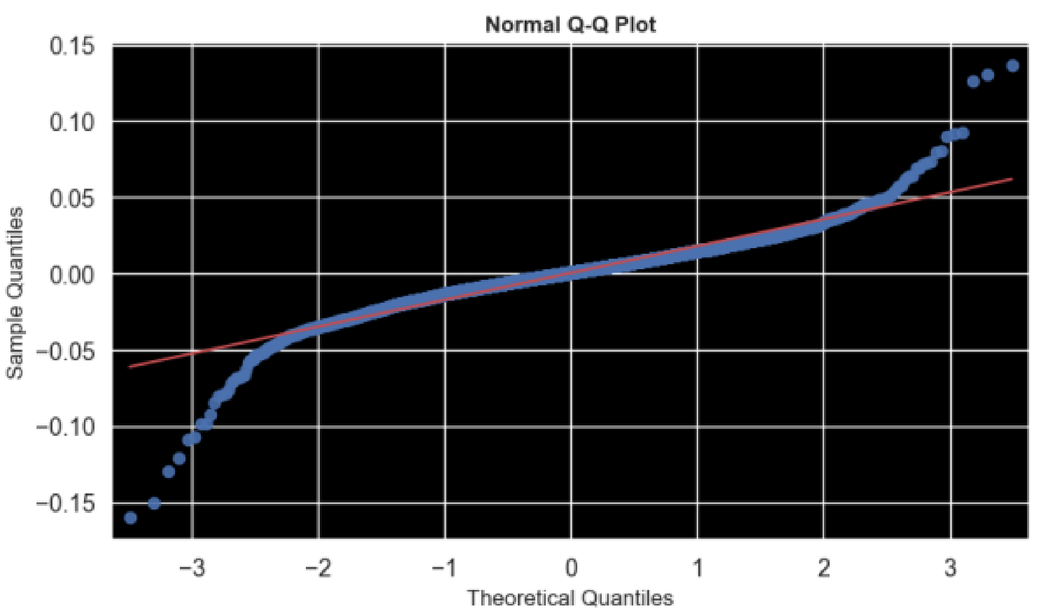
Figura 3.7 - Distribuição dos retornos do índice IBOVESPA de 2005 até 2021



Fonte - Elaboração própria com base nos dados do Yahoo Finance

Visualmente, o histograma parece mostrar uma distribuição gaussiana para os retornos do índice, no entanto, vale ressaltar que a amostra é grande e devemos realizar um teste de normalidade para conferir se de fato a distribuição probabilística dos retornos é normal. Através do teste de normalidade de Shapiro-Wilk e a plotagem do gráfico quantil-quantil (Figura 3.8) poderemos confirmar a hipótese.

Figura 3.8 - Gráfico quantil-quantil para teste de normalidade. O teste de Shapiro-Wilk revela um p-valor de 1,3 * 10^-41, dessa forma, rejeitando a hipótese nula de que a distribuição é normal



Fonte - Elaboração própria com base nos dados do Yahoo Finance

Embora a distribuição não é normal, devido ao tamanho da mostra, é possível tolerar certo nível de assimetria e tomar a suposição que a distribuição é gaussiana para diversas aplicações de testes e modelos estatísticos. Outros parâmetros a serem considerados nos estudos do formato da distribuição é a curtoses e assimetria.

A assimetria, ou skweness em inglês, é usado para descrever a simetria da coleção de dados, uma distribuição perfeitamente simétrica dos dados teria 0 como valor da assimetria. Valores positivos significam que a distribuição dos dados possui uma calda direita maior que a esquerda, e valores negativos de assimetria representam o oposto, a calda da esquerda maior do que a da direita. A assimetria dos retornos do índice é –0.44. Segundo Bill Mcneese, valores de assimetria entre –0,5 e 0,5 representam uma distribuição majoritariamente simétrica. A equação de assimetria é dada por:

$$S(R) = \frac{E\left[(R - E(R))^3\right]}{\sigma_R^3}$$

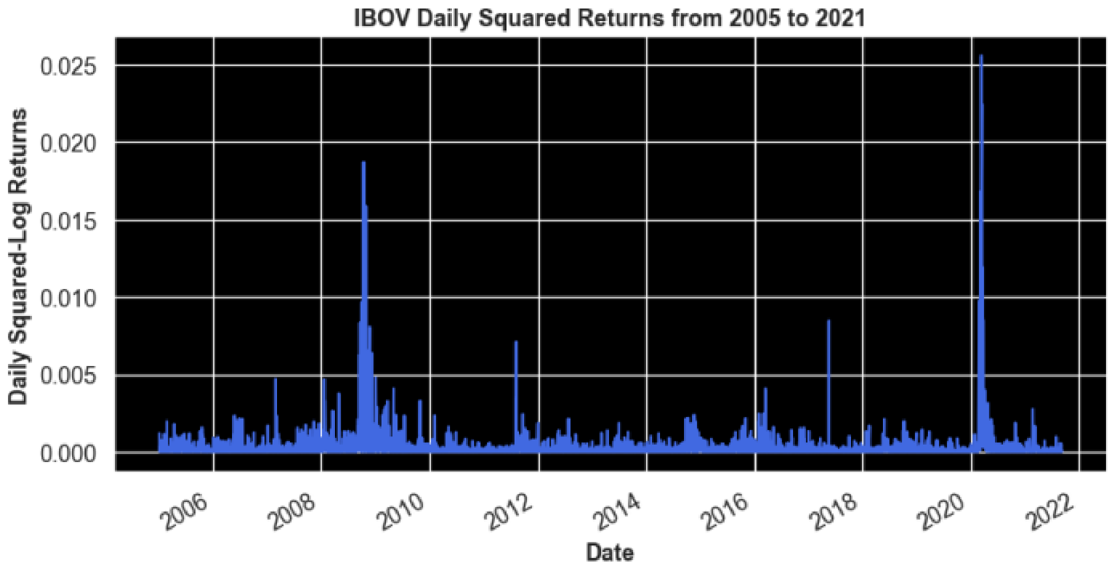
Outra métrica que descreve o formato da distribuição é a curtoses, no mesmo artigo de Bill Mcneese, é citado a fala do Dr. Wheeler, crítico ao uso da assimetria e curtoses em processos estatísticos. “O parâmetro curtoses é uma medida de pesos combinados da calda relativa ao resto da distribuição”. Curtoses negativa evidencia uma calda suave em relação ao resto da distribuição, enquanto um valor positivo desse parâmetro demonstra uma calda mais evidente em contraste com outras regiões da distribuição. O valor da curtose de uma distribuição perfeitamente normal é 3, muitas vezes é conveniente calcular o excesso de curtoses, subtraindo o valor por 3 e comparando com 0. O excesso de curtose da distribuição dos retornos do índice é 9,064, apresentando caldas bem evidentes em comparação com o centro da distribuição. A equação da curtose é dada por:

$$S(R) = \frac{E\left[(R - E(R))^4\right]}{\sigma_R^4}$$

Embora a assimetria da distribuição dos dados do IBOVESPA é praticamente a de uma distribuição normal, simétrica, o que pode levar ao engano quando apenas o histograma é visualizado, a curtoses está bem distante de uma distribuição normal, convergindo para o resultado de teste de normalidade.

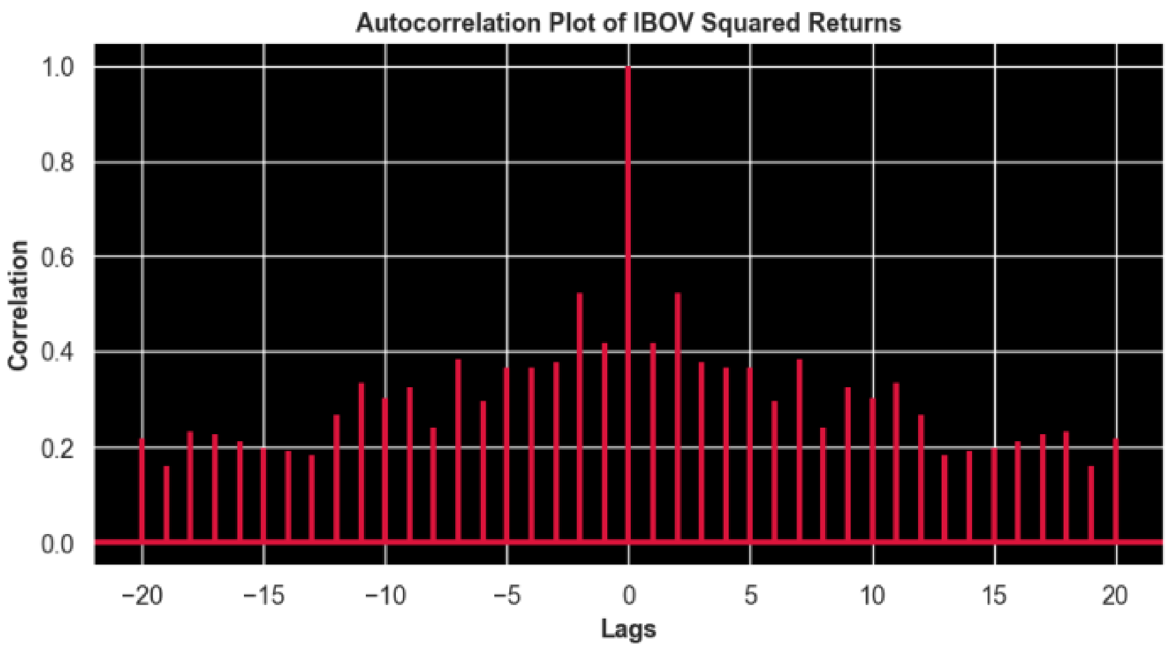
Elevando ao quadrado o log dos retornos dos índices, temos os seguintes gráficos de séries ilustrados pela figura 3.9 e 3.10:

Figura 3.9 - Retornos diários ao quadrado do índice IBOVESPA



Fonte - Elaboração própria com base nos dados do Yahoo Finance

Figura 3.10 - Auto correlação dos retornos diários ao quadrado do índice IBOVESPA



Fonte - Elaboração própria com base nos dados do Yahoo Finance

Podemos perceber que elevando os retornos do índice ao quadrado, algumas propriedades de séries temporais estacionárias não são mais encontradas. Ao começar pela variância constante ao longo do intervalo na figura 3.8. Devido ao expoente quadrático, os valores apresentam um gap maior entre si, e uma variância menos constante ao longo da série. Os retornos também não parecem aleatoriamente distribuído ao redor do valor zero, uma vez que a sequência só apresenta valores positivos.

No correlograma apresentado pela figura 3.8 as auto-correlações, embora simétricas em relação ao lag de valor zero, estão distantes de serem nula. Também é possível notar uma tendência decrescente ao passo que o tempo se distância do marco 0, essas características invalidam as propriedades de estacionariedade fraca de séries.

4. Referências

PRADO, R.; WEST, M. **Time Series Modelling, Inference and Forecasting**. University of California, Santa Cruz.

MCNEESE, Bill. Are the Skewness and Kurtosis Useful Statistics? **SPC Excel**, 2016. Disponível em: https://www.spcforexcel.com/knowledge/basic-statistics/are-skewness-and-kurtosis-useful-statistics