



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA
DEPARTAMENTO DE MATEMÁTICA APLICADA



MARIANA AIKO ICHICAWA OGIDO

Análise de Séries Temporais - Modelos Autorregressivos

Campinas
28/06/2023

MARIANA AIKO ICHICAWA OGIDO

Análise de Séries Temporais - Modelos Autorregressivos

Monografia apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos para obtenção de créditos na disciplina Projeto de Extensão Supervisionado, sob a orientação do(a) Prof. Lúcio Tunes dos Santos.

Conteúdo

| | | |
|----------|---|-----------|
| 1 | Introdução | 4 |
| 2 | Fundamentos | 4 |
| 2.1 | O que é uma série temporal? | 4 |
| 2.2 | Tipos de séries | 4 |
| 2.3 | Como analisar uma série temporal | 4 |
| 2.4 | Características de séries temporais financeiras | 5 |
| 2.5 | A matemática das séries temporais | 5 |
| 2.5.1 | Discreta ou Contínua | 6 |
| 2.5.2 | Univariada ou Multivariada | 6 |
| 2.5.3 | Unidimensional ou Multidimensional | 6 |
| 2.6 | Processos Estocásticos | 6 |
| 2.7 | Processos Estacionários | 8 |
| 3 | Modelos para Séries Temporais | 9 |
| 3.1 | Notação | 9 |
| 3.2 | Modelos lineares estacionários | 10 |
| 3.2.1 | Processo linear geral | 10 |
| 3.2.2 | Condições de estacionariedade e invertibilidade | 11 |
| 3.3 | Modelos de auto-regressão (AR) | 11 |
| 3.3.1 | Simulação AR(1) | 12 |
| 3.3.2 | Características do processo AR(p) | 12 |
| 3.4 | Modelo de Médias Móveis (MA) | 13 |
| 3.4.1 | Simulação MA(1) | 13 |
| 3.4.2 | Características do processo MA(q) | 14 |
| 4 | Experimentos Numéricos | 14 |
| 4.1 | Obtenção e tratamento dos dados | 15 |
| 4.2 | Análise da autocorrelação histórica | 16 |
| 4.3 | Conclusões | 19 |

1 Introdução

A análise de séries temporais pode ser utilizada em uma ampla gama de áreas com o objetivo de realizar projeções. Algumas das diferentes áreas de estudo de séries temporais são as finanças quantitativas, meteorologia, e saúde. Além de ser usada para projeções, pode ser utilizada para reconhecimento de padrões, detecção de anomalias, e classificação, por exemplo.

O objetivo deste trabalho é realizar um estudo em torno de séries temporais até a introdução de alguns dos modelos mais conhecidos para previsões. Além disso, o trabalho tem como foco a análise de séries temporais financeiras.

2 Fundamentos

A seguir serão apresentadas definições relevantes e aspectos importantes sobre séries temporais de forma a construir o conhecimento desde o básico do básico até a introdução de modelos autorregressivos (Morettin and Tolo [2006]).

2.1 O que é uma série temporal?

Uma série temporal é um conjunto de observações ordenadas no tempo. Podemos observar diversas séries temporais como a média mensal da temperatura de Barão Geraldo, o preço de fechamento diário de ações da Petrobras, e registro de marés no porto de Santos.

2.2 Tipos de séries

O primeiro e segundo exemplos são séries temporais discretas, enquanto que o terceiro exemplo é uma série temporal contínua.

2.3 Como analisar uma série temporal

De maneira geral, para analisar uma série precisamos que ela seja discreta. Assim, quando temos uma série contínua, ela deve ser amostrada em intervalos de tempo de mesmo tamanho de modo a ser convertida para uma série discreta.

Existem dois tipos de análises que podem ser realizadas nas séries temporais. O primeiro tipo é feito no domínio temporal e os modelos propostos são modelos paramétricos, que possuem um número finito de parâmetros. Já o segundo tipo de análise é realizada no domínio de frequências e os modelos propostos são os modelos não-paramétricos.

2.4 Características de séries temporais financeiras

As séries temporais financeiras apresentam algumas características interessantes, podendo ser também observadas em séries temporais de outras naturezas, dentre elas temos tendências, sazonalidade, pontos influentes ou atípicos - comumente chamado de *outliers*, heteroscedasticidade condicional, e não-linearidade.

Além disso, os retornos financeiros apresentam alguns fatos estilizados, ou seja uma aproximação teórica para o que é observado empiricamente.

1. Os retornos em geral são não-auto-correlacionados;
2. os quadrados dos retornos são auto-correlacionados, apresentando uma correlação de defasagem um pequena e depois uma queda lenta das demais;
3. séries de retornos apresentam agrupamentos de volatilidade ao longo do tempo;
4. a distribuição (incondicional) dos retornos apresenta caudas mais pesadas - também chamado de *fat tails* - que uma distribuição normal; além disso, a distribuição, embora aproximadamente simétrica, é em geral leptocúrtica;
5. algumas séries de retornos são não-lineares. (Morettin and Tolo [2006])

2.5 A matemática das séries temporais

Uma série temporal é um conjunto de observações ordenadas no tempo. Podemos representar matematicamente, uma série por um vetor $\mathbf{Z}(\mathbf{t})$ de ordem $r \times 1$ onde \mathbf{t} é um vetor $p \times 1$.

Podemos classificar as séries de três formas, se ela é discreta ou contínua, univariada ou multivariada, unidimensional ou multidimensional. A seguir, são apresentadas as características de cada uma dessas classificações e alguns exemplos de séries temporais.

2.5.1 Discreta ou Contínua

Uma série discreta é referente a observações espaçadas no tempo, enquanto que uma série contínua diz respeito a observações contínuas. Uma série discreta é, por exemplo, o PIB anual do Brasil. Já a série temporal contínua, podemos observá-la, por exemplo, em fenômenos da natureza como a temperatura.

2.5.2 Univariada ou Multivariada

A classificação da série em univariada ou multivariada é referente a quantidade de séries que podemos gerar com uma única fonte de observação. Um exemplo de série univariada pode ser o número de casos de Covid-19 por mês no Brasil. E um exemplo de série multivariada pode ser medidas de pressão uterina e pressão sanguínea em uma mulher durante o parto.

$$Z(t) = [\text{Casos de Covid-19(mês)}]$$
$$Z(t) = [\text{Pressão Sanguínea}(t), \text{Pressão Uterina}(t)]$$

2.5.3 Unidimensional ou Multidimensional

Já a classificação em unidimensional ou multidimensional, diz quanto às variáveis dependentes da série observada. Uma série temporal tem como variável dependente o parâmetro t , mas não é necessariamente exclusivo. A série temporal pode ter outros parâmetros físicos como o espaço ou volume. Um exemplo de série unidimensional pode ser o número de nascimentos no mundo. Uma série multidimensional pode ser, por exemplo, a pressão atmosférica de um mergulhador no mar, neste caso o dado observado depende de um vetor $t = [\text{tempo}, \text{profundidade}]$.

$$Z(t) = [\text{Pressão sobre o mergulhador (tempo, profundidade)}]$$

2.6 Processos Estocásticos

Um processo estocástico refere-se a uma série de eventos aleatórios em que cada um desses eventos possui uma distribuição de probabilidade. De maneira mais formal podemos definir um processo estocástico da seguinte forma.

Definição 1 *Seja T um conjunto arbitrário definido em um espaço de probabilidade. Um processo estocástico é uma família*

$$Z = \{Z(t), t \in T\}, \quad (1)$$

tal que, para cada $t \in T$, $Z(t)$ é uma variável aleatória.

O conjunto T é normalmente tomado como o conjunto dos inteiros ou reais e representa a passagem do tempo. Além disso, cada $t \in T$, $Z(t)$ é uma variável aleatória.

Ademais, especificamos um processo estocástico da seguinte forma.

Sejam t_1, t_2, \dots, t_n elementos de T , e

$$F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z(t_1) \leq z_1, \dots, Z(t_n) \leq z_n\}, \quad (2)$$

a função de distribuição cumulativa de Z_1, \dots, Z_n ou $Z(t_1), \dots, Z(t_n)$, onde a equação (2) representa a probabilidade de que a variável aleatória $Z(t_i)$ assumira um valor menor ou igual z_i , $i = 1, \dots, n$.

O processo estocástico $Z = \{Z(t), t \in T\}$ estará especificado se conhecermos as distribuições das variáveis aleatórias $Z(t_1), \dots, Z(t_n)$, $t_1, \dots, t_n \in T$.

A função densidade de probabilidade correspondente a distribuição cumulativa é a f.d.p. $f(z_1, \dots, z_n; t_1, \dots, t_n)$. E a média da v.a. $Z(t)$ é dada por,

$$\mu(t) = E\{Z(t)\} = \int_{-\infty}^{\infty} z f(z; t) dz, \quad (3)$$

onde $t \in T$.

Já a função de autocovariância (facv) de $Z(t_1, t_2)$ é

$$\gamma(t_1, t_2) = Cov\{Z(t_1), Z(t_2)\} = E\{Z(t_1)Z(t_2)\} - E\{Z(t_1)\}E\{Z(t_2)\}, \quad (4)$$

com $t_1, t_2, \in T$.

No caso particular em que $t_1 = t_2 = t$ temos a seguinte função de autocovariância:

$$\gamma(t, t) = Var\{Z(t)\} = V(t) = E\{Z^2(t)\} - E^2\{Z(t)\}. \quad (5)$$

A função de autocorrelação do processo é definida por

$$\rho_\tau = \frac{\gamma(t_1, t_1 - \tau)}{\gamma(t_1, t_1)}, \quad (6)$$

onde $\tau = |t_1 - t_2| \in T, t_1 > t_2$.

Dentre os processos estocásticos, podemos distinguí-los de acordo com três características,

1. independência relativa à origem dos tempos - processo estacionário ou não estacionário;
2. função densidade de probabilidade que caracteriza o processo - processo normal ou não normal;
3. independência do processo de valores em instantes precedentes - processo Markoviano ou não Markoviano.

2.7 Processos Estacionários

De maneira informal, um processo Z é estacionário se a escolha de uma origem de tempos não é relevante e não afeta suas características estatísticas. Ou seja, as características estatísticas como a média e variância são preservadas independente do período analisado da série.

Em geral, as diversas formas de ruídos podem ser consideradas processos estacionários. Para fins de análise, existem duas formas de estacionariedade, a estacionariedade fraca e a estrita.

Definição 2 *Um processo estocástico $Z = \{Z(t), t \in T\}$ é dito estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações do tempo, ou seja,*

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n) \quad (7)$$

para quaisquer $t_1, \dots, t_n, \tau \in T$.

Como consequência da equação acima, pode-se mostrar que a média $\mu(t)$ e a variância $V(t)$ são constantes, para todo $t \in T$.

Definição 3 *Um processo estocástico $Z=\{Z(t), t \in T\}$ é dito fracamente estacionário se e somente se*

1. $E\{Z(t)\}$ é constante para todo $t \in T$;
2. $E\{Z^2(t)\} < \infty$, para todo $t \in T$;
3. $\gamma(t_1, t_2)$ é uma função de $|t_1 - t_2|$.

3 Modelos para Séries Temporais

Os modelos para séries temporais podem ser classificados em duas classes de acordo com o número de parâmetros utilizados. Os modelos paramétricos são aqueles os quais o número de parâmetros é finito, e os modelos não paramétricos contam com um número infinito de parâmetros. Neste estudo, trabalharemos apenas com modelos paramétricos lineares.

Seguindo a metodologia de Box and Jenkins [1976] para modelos paramétricos, a estratégia para a construção de um modelo é feita a partir do seguinte ciclo iterativo:

1. **especificação** de uma classe geral de modelos a ser considerada para análise;
2. **identificação** de um modelo, com base em algumas análises como a de autocorrelações e autocorrelações parciais;
3. **estimação** dos parâmetros do modelo;
4. **verificação** ou **diagnóstico** do modelo ajustado através de uma análise de resíduos, para saber se o modelo é adequado para seu objetivo.

3.1 Notação

A seguir definiremos alguns operadores a fim de facilitar a manipulação dos modelos que serão apresentados.

1. Operador B de translação para o passado:

$$BZ_t = Z_{t-1},$$

$$B^m Z_t = Z_{t-m};$$

2. operador F de translação para o futuro:

$$FZ_t = Z_{t+1},$$

$$F^m Z_t = Z_{t+m};$$

3. operador Δ de diferença:

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t,$$

$$\Delta = 1 - B;$$

4. operador S de soma:

$$SZ_t = \sum_{j=0}^{\infty} Z_{t-j} = Z_t + Z_{t-1} + \dots = (1 + B + B^2 + \dots)Z_t,$$

$$S = \Delta^{-1}.$$

De maneira geral, chamaremos de *lag* quando estivermos falando de uma série transladada para o passado.

3.2 Modelos lineares estacionários

3.2.1 Processo linear geral

O modelo de filtro linear supõe que a série temporal é gerada através de um filtro linear e tem a seguinte forma:

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = E(Z_t) + 1 + \psi_1 B + \psi_2 B^2 + \dots, \quad (8)$$

onde a_t é um ruído branco e ψ_1, ψ_2, \dots é uma sequência de parâmetros em que

$$\sum_{k=0}^{\infty} \psi_k^2 < \infty.$$

Além disso, podemos definir

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots \quad (9)$$

denominada função de transferência do filtro.

Seja,

$$\tilde{Z}_t = Z_t - \mu, \quad (10)$$

podemos escrever \tilde{Z}_t em função de seus valores passados adicionado a um ruído

a_t :

$$\tilde{Z}_t = \pi_1 \tilde{Z}_{t-1} + \pi_2 \tilde{Z}_{t-2} + \dots + a_t = \sum_{j=1}^{\infty} \pi_j \tilde{Z}_{t-j} + a_t \quad (11)$$

3.2.2 Condições de estacionariedade e invertibilidade

Proposição 1 *Um processo linear será considerado estacionário se a série $\psi(B)$ convergir para $|B| \leq 1$.*

Proposição 2 *Um processo linear será invertível se a série $\pi(B)$ convergir para $|B| \leq 1$.*

3.3 Modelos de auto-regressão (AR)

Para o caso em que para a equação (11) temos $\pi_j = 0, j > p$, obtemos um modelo auto-regressivo de ordem p , denotado por $AR(p)$

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t, \quad (12)$$

os pesos π_j foram renomeados para ϕ_j de acordo com a notação usual.

O caso mais simples do modelo auto-regressivo é de ordem $p = 1$, $AR(1)$

$$\tilde{Z}_t = \phi \tilde{Z}_{t-1} + a_t. \quad (13)$$

Observe que \tilde{Z}_t depende apenas de \tilde{Z}_{t-1} e do ruído no instante t .

3.3.1 Simulação AR(1)

A seguir será simulado uma série de 50 observações com modelo AR(1)

$$Z_t = 0.7Z_{t-1} + a_t$$

onde $a_t \sim N(0, 1)$.

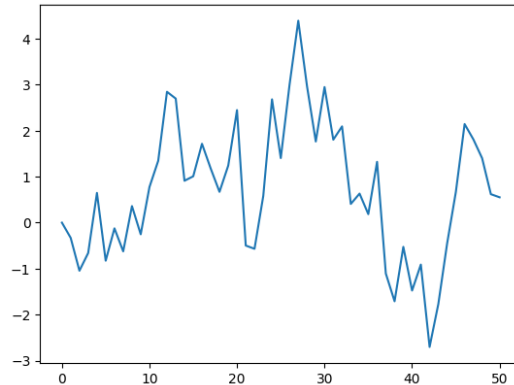


Figura 1: Gráfico da série simulada AR(1): $Z_t = 0.7Z_{t-1} + a_t$.

3.3.2 Características do processo AR(p)

Temos que $\pi(B) = \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ é finito, então não há restrição dos parâmetros para garantir a invertibilidade de Z_t . Já a condição de estacionariedade é de que as raízes $G_i^{-1}, i = 1, \dots, p$ da equação $\phi(B) = 0$ estejam fora do círculo unitário.

Quanto a função de autocorrelação, temos duas possibilidades de comportamento:

1. se G_i for real, a função tem a forma de um amortecimento exponencial;
2. um par de raízes complexas conjugadas contribui com um termo na forma de uma senóide amortecida.

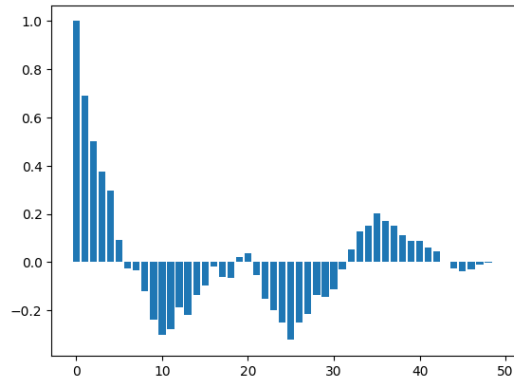


Figura 2: Função de autocorrelação amostral da série simulada AR(1): $Z_t = 0.7Z_{t-1} + a_t$.

3.4 Modelo de Médias Móveis (MA)

Seja um processo linear com $\phi_j = 0, j > q$, temos um processo de médias móveis de ordem q , denotado por MA(q), do inglês *Moving Average*, com a seguinte notação

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (14)$$

foi feita a seguinte substituição de $\phi_j = -\theta_j$.

O modelo mais simples é o de ordem $q = 1$, o MA(1),

$$\tilde{Z}_t = a_t - \theta a_{t-1} \quad (15)$$

3.4.1 Simulação MA(1)

Considere o processo de 50 observações gerado pelo seguinte modelo MA(1)

$$Z_t = a_t - 0.7a_{t-1},$$

onde $a_t \sim N(0, 1)$.

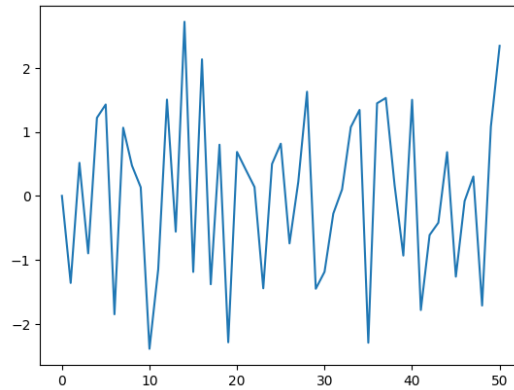


Figura 3: Gráfico da série simulada MA(1): $Z_t = a_t - 0.7a_{t-1}$.

3.4.2 Características do processo MA(q)

Neste processo, sabemos que $\psi(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, então $\psi(B)$ sempre converge para $|B| \leq 1$. Logo, o processo é estacionário sem restrições sobre os parâmetros θ_j . Já para a invertibilidade, as raízes de $\theta(B) = 0$ devem estar fora do círculo unitário.

Em um processo MA(q), a função de autocorrelação é zero para *lags* maiores que q .

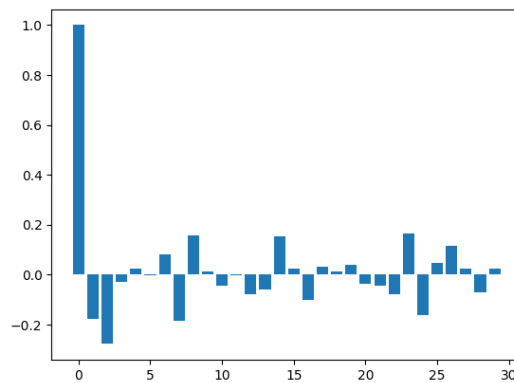


Figura 4: Função de autocorrelação amostral da série simulada MA(1): $Z_t = a_t - 0.7a_{t-1}$.

4 Experimentos Numéricos

Analisando a função de autocorrelação de uma série em intervalos móveis, podemos entender o comportamento da série ao longo do tempo. No mercado financeiro, existem diferentes regimes e estados do mercado, os quais variam com o tempo, ou seja,

em alguns momentos, dentro de um intervalo móvel definido, a série pode apresentar um comportamento em que os dados seguem uma tendência, podendo ser representado por um modelo AR, ou um comportamento de reversão a média, podendo ser representado por um modelo MA. Entende-se que se uma série possui uma alta autocorrelação, seja um lag k , o valor da série em um t no tempo, segue um comportamento parecido com o valor da série em $t - k$.

4.1 Obtenção e tratamento dos dados

Para o experimento, utilizaremos a série histórica diária do Índice Ibovespa em um período de 5 anos, entre 26/06/2018 a 23/06/2018. Os dados estão disponíveis no Yahoo Finance.

| Date | Adj Close |
|------------|-----------|
| 06/26/2018 | 71405 |
| 06/27/2018 | 70609 |
| 06/28/2018 | 71767 |
| 06/29/2018 | 72763 |
| 07/02/2018 | 72840 |

Tabela 1: Primeiros cinco dados do fechamento ajustado da série histórica dos últimos cinco anos do Índice Ibovespa.



Figura 5: Preço de fechamento ajustado do Índice Ibovespa dos últimos cinco anos.

É claro que a série histórica de fechamento ajustado não apresenta estacionariedade. Para iniciarmos a análise é necessário tornar a série estacionária. Nesse caso, calcularemos os retornos diários percentuais, visando obter uma série estacionária, pela seguinte fórmula

$$r_i = \frac{c_i - c_{i-1}}{c_{i-1}} = \frac{c_i}{c_{i-1}} - 1, \quad (16)$$

onde c_i é o preço de fechamento ajustado no dia i e r_i é o retorno percentual diário no dia i .

Traçando o gráfico dos retornos diários, obtemos a figura 6, com as características de média igual a 0.0006 e variância igual a 0.0003.

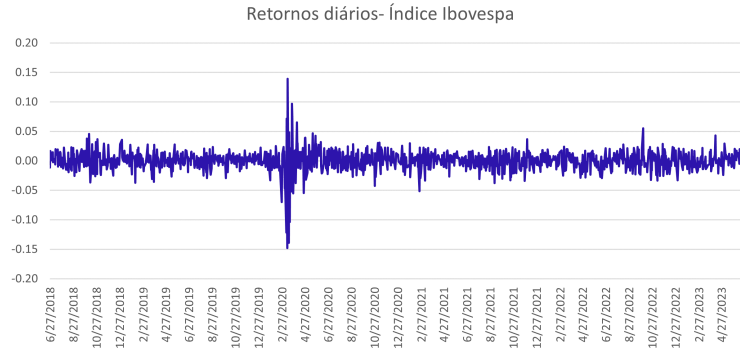


Figura 6: Retorno percentual diário do Índice Ibovespa dos últimos cinco anos.

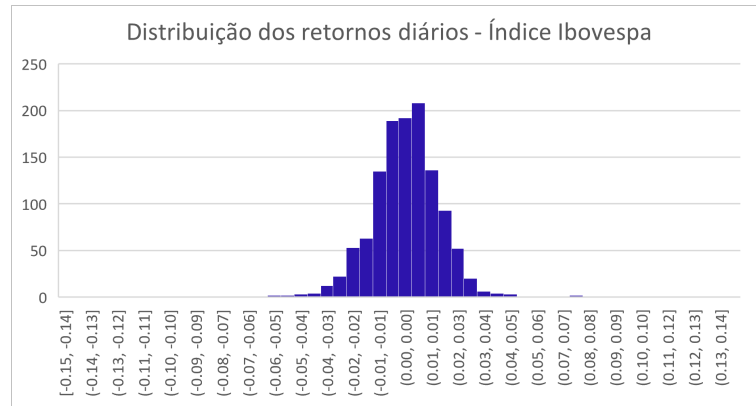


Figura 7: Distribuição de retornos diários do Índice Ibovespa dos últimos cinco anos.

4.2 Análise da autocorrelação histórica

Foram analisadas as correlações históricas móveis, em uma janela de 70 dias úteis escolhida arbitrariamente, para diferentes *lags*. Os *lags* escolhidos foram de 1, 5, 10, 20, 60. Os valores foram escolhidos, pois na série do Índice Ibovespa representam 1 dia, 1 semana, 2 semanas, 1 mês e 3 meses, respectivamente, que são intervalos relevantes para dados desse tipo.

Na escolha de um *lag* de 1 dia, podemos observar que do início da série até meados de abril de 2022, os retornos apresentavam uma maior autocorrelação negativa. De forma simples, o retorno em um dia, em geral, era no sentido contrário ao retorno do dia anterior. Em seguida, houve um período em que as autocorrelações se mantiveram positivas, ou seja, o retorno em um dia, em geral, ocorria no mesmo sentido do retorno do dia anterior. Já ao final da série, não é muito claro o sentido que os retornos podem tomar, apesar das autocorrelações estarem mais positivas do que negativas.

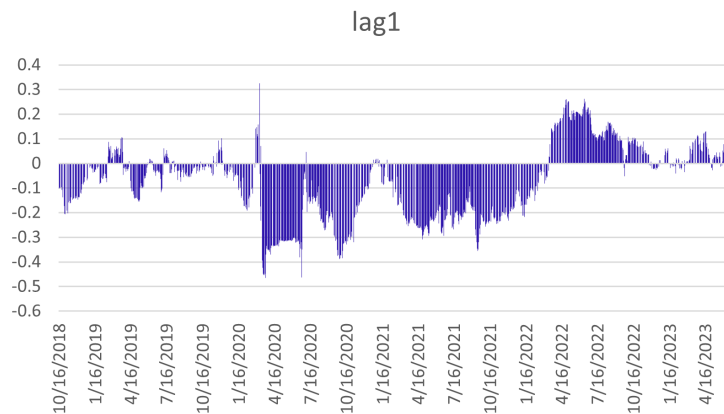


Figura 8: Autocorrelação histórica em uma janela de 70 pontos e *lag* 1.

A autocorrelação histórica de um lag de 5 pontos apresentou uma grande variabilidade nos períodos em que era mais negativa ou mais positiva, principalmente entre 2021 e 2022.

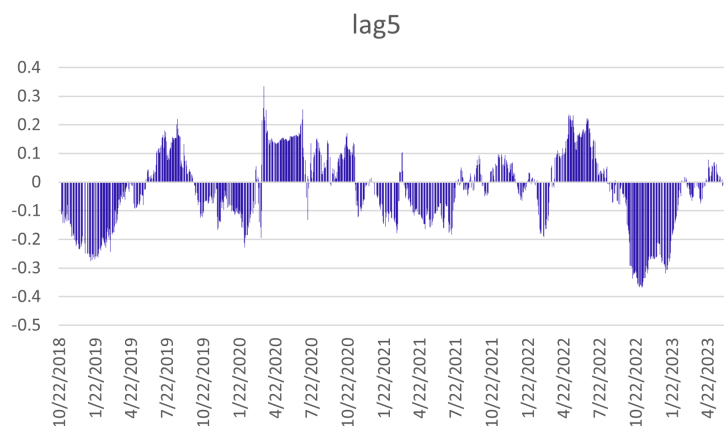


Figura 9: Autocorrelação histórica em uma janela de 70 pontos e *lag* 5.

Já para um *lag* de 10 pontos, não apresentam nenhum grande período de

autocorrelação positiva ou negativa.

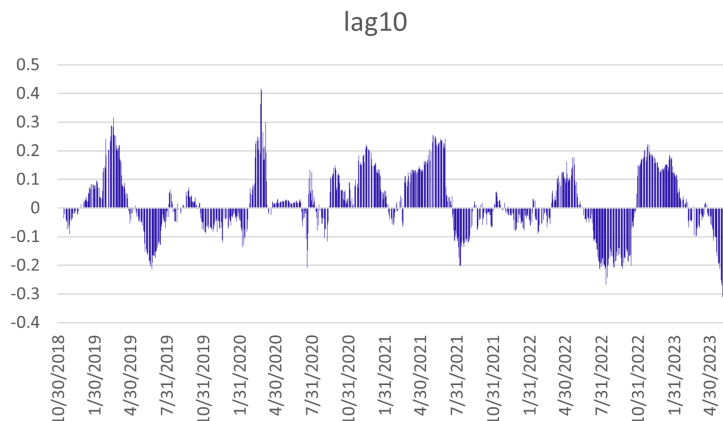


Figura 10: Autocorrelação histórica em uma janela de 70 pontos e lag 10.

Para um lag de 20 pontos, as autocorrelações se mostraram mais presentes para o lado negativo dentro do período observado.

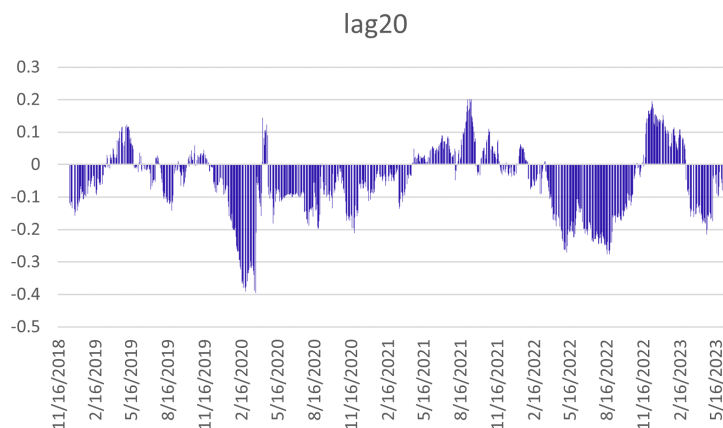


Figura 11: Autocorrelação histórica em uma janela de 70 pontos e lag 20.

Utilizando um lag de 60 pontos, observamos períodos mais longos em que a correlação se mantém com o mesmo sinal. Além disso, existem alguns períodos em que os retornos são praticamente decorrelacionados (entre o começo de junho até o começo de novembro de 2021), ou seja, nada se pode dizer sobre os retornos desse período com base em retornos passados.

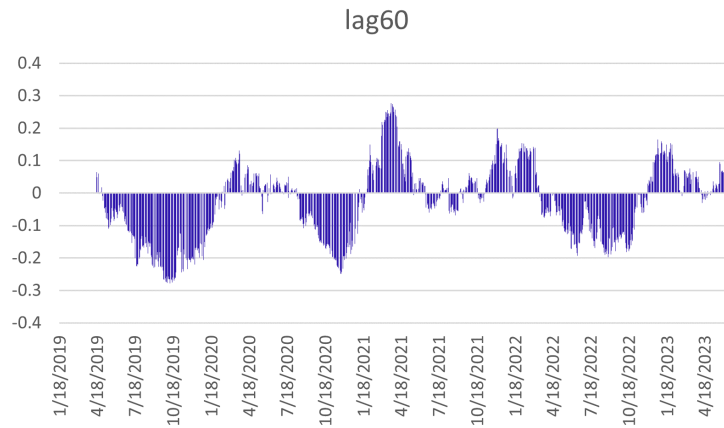


Figura 12: Autocorrelação histórica em uma janela de 70 pontos e lag 60.

Na figura 13, temos um gráfico mostrando a autocorrelação histórica média para vários valores de lag . Nesse gráfico podemos observar que dependendo do lag escolhido, podemos ter uma autocorrelação mais forte positivamente ou negativamente. Por exemplo para lags de 1, 7, 13 e 20, temos que a média da autocorrelação é mais forte para algum dos lados.

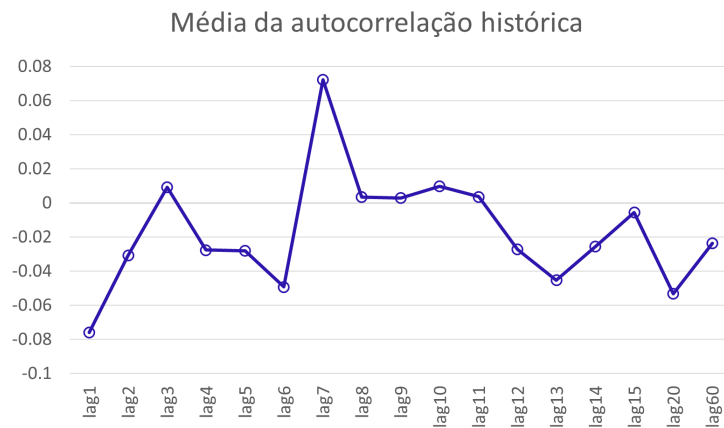


Figura 13: Autocorrelação média para vários valores de lag

4.3 Conclusões

Realizando uma análise da autocorrelação histórica para diferentes $lags$ podemos entender o comportamento e a direção dos retornos em certos períodos. Essa informação é de grande valor para escolher e definir estratégias de investimento condizentes com o comportamento do mercado financeiro no período vigente.

Referências

Yahoo finance. <https://finance.yahoo.com/>, June 2023.

George E. P. Box and Gwilym M. Jenkins. *Box and Jenkins: time series analysis, forecasting and control*. Holden-DAY, Incorporated, 1976.

Pedro A. Morettin and Clélia M. C. Tolo. *Análise de Séries Temporais*. ABE - Projeto Fisher, 2006.

Sheldon Ross. *A first course in probability*. Pearson, 2010.