

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

U LISBOA

UNIVERSIDADE
DE LISBOA

O impacto da inflação no S&P 500

Beatriz Alexandra Monteiro Barroso

Mestrado em Matemática Financeira

Orientadores:

Professora Doutora, Diana Elisabeta Aldea Mendes, Professora
Associada

outubro, 2023

Departamento de Finanças

Departamento de Matemática

O impacto da inflação no S&P 500

Beatriz Alexandra Monteiro Barroso

Mestrado em Matemática Financeira

Orientadores:

Professora Doutora, Diana Elisabeta Aldea Mendes, Professora
Associada

ISCTE-IUL

outubro, 2023

“Foi o tempo que dedicaste à tua rosa que a fez tão importante.”

Antoine de Saint-Exupéry em O Príncipezinho

Agradecimentos

Em primeiro lugar, gostaria de agradecer à minha orientadora Diana Mendes por me ter disponibilizado do seu pouco tempo para me acompanhar neste projeto e me ter dado a força e a coragem em setembro de 2022 no gabinete 207 para não desistir.

Em seguida, quero agradecer à minha família, que me acompanhou naquele que foi o ano em que enfrentei mais desafios em termos profissionais, acadêmicos e pessoais. Que me soube dar o apoio incondicional e essencial para alcançar os meus objetivos.

Não por último menos importantes, agradecer aos meus amigos que me deram sempre o ombro que eu precisava para me apoiar.

Sem os pilares da minha vida acima listados, teria certamente desistido sem antes tentar.

Resumo

O presente trabalho tem como principal objetivo o estudo do impacto da inflação no índice S&P 500 e identificar o método que apresenta melhor adaptação à previsão do índice S&P 500 em contexto económico e por isso identificar a correlação que existe entre as variáveis que são influenciadas pelo fenómeno.

A resposta dos bancos centrais à inflação é importante e incisiva. Se o Banco Central adotar uma abordagem agressiva para controlar a inflação, aumentando significativamente as taxas de juros, poderá levar a criar pressão de baixa sobre o mercado de ações, incluindo o S&P 500.

Embora as ações tenham a capacidade de superar a inflação a longo prazo, períodos de alta inflação podem afetar negativamente o investimento nos mercados financeiros. Além disso, a inflação pode levar a aumentos nas taxas de juros, o que pode tornar os investimentos em renda fixa mais atraentes em relação às ações.

Neste estudo, as variáveis utilizadas foram o S&P 500, o Consumer Price Index, Real Gross Domestic Product and Unemployment Rate, todas com frequência trimestral e 60 observações de janeiro de 2015 a dezembro de 2022.

Os modelos univariados de séries temporais são usados para analisar e prever o comportamento de uma única série temporal ao longo do tempo e aqui tornam-se importantes uma vez que se pretende verificar a correlação não-linear existente entre a variável S&P 500 e as demais. Assim, é necessário o estudo individual do melhor modelo a aplicar.

Palavras-chave: U.S.; Inflação; S&P 500; Consumer Price Index; Gross Domestic Product; Unemployment Rate; Série temporal; Previsão; Modelos ARIMA; Modelos VAR;

Classificação JEL: N12; E31; C43; E24; C22; C32; C50; E27

Abstract

The main objective of this work is to study the impact of inflation on the S&P 500 index and identify the method that best adapts to the prediction of the S&P 500 index in an economic context and therefore identify the correlation that exists between the variables that are influenced by the phenomenon.

Central banks' response to inflation is important and incisive. If the Central Bank takes an aggressive approach to controlling inflation by significantly raising interest rates, it could lead to creating downward pressure on the stock market, including the S&P 500.

Although stocks can outperform inflation over the long term, periods of high inflation can negatively affect investment in financial markets. Additionally, inflation can lead to increases in interest rates, which can make fixed income investments more attractive relative to stocks.

In this study, the variables used were the S&P 500, the Consumer Price Index, Real Gross Domestic Product and Unemployment Rate, all with a quarterly frequency and 60 observations from January 2015 to December 2022.

Univariate time series models are used to analyze and predict the behavior of a single time series over time and here they become important since the aim is to verify the non-linear correlation that exists between the S&P 500 variable and the others. Therefore, it is necessary to individually study the best model to apply.

Keyword: U.S.; Inflation; Unemployment; Time-Series Models; Econometric Modeling; Forecasting and Simulation: Models and Applications

JEL Classification: N12; E31; E24; C22; C5; E27

Índice

CAPÍTULO 1.....	3
1.1. Enquadramento teórico	3
1.1.1. O conceito de inflação	3
1.1.2. A evolução da inflação nos Estados Unidos da América	4
1.1.3. O índice S&P 500	5
1.1.4. O impacto da inflação no S&P 500	7
CAPÍTULO 2.....	9
2.1. Séries Temporais.....	9
2.2. Descrição das séries temporais	10
2.2.1. Assimetria	10
2.3. Estacionaridade	11
2.3.1. Diferenças Sucessivas	12
2.3.2. Diferenças Logaritmizadas.....	12
2.3.3. Ruído Branco.....	12
2.3.4. Testes de Raiz Unitária	13
2.4. Modelos Univariados.....	14
2.4.1. Modelos Autorregressivos - $AR(p)$	15
2.4.2. Modelos de Média Móvel - $MA(q)$	16
2.4.3. Modelos Autorregressivos de Médias Móveis – $ARMA(p, q)$	16
2.4.4. Modelos Autorregressivos Integrados de Médias Móveis - $ARIMA(p, d, q)$	17
2.4.5. A Metodologia de Box-Jenkins	17
2.4.6. Modelos ARCH/GARCH	18
2.4.7. O modelo GARCH.....	19
2.5. Previsão.....	20
2.5.1. Erro de Previsão.....	20
CAPÍTULO 3.....	22
3.1. Análise das Séries Temporais	22
3.1.1. Estacionaridade	25
3.2. Estimação do Modelo	26
3.2.1. Estimação dos Modelos Univariados.....	26
3.2.1.1. Análise dos Resíduos para os Modelos Univariados	29
3.2.1.2. Previsão das variáveis.....	30
3.2.2. Estimação dos Modelos Multivariados.....	31
CAPÍTULO 4.....	33
CAPÍTULO 5.....	35

Introdução

A inflação é um conceito económico fundamental descrito pelo aumento generalizado e contínuo dos preços dos bens e serviços numa economia durante um longo período. Tem um impacto profundo em diversos aspetos da economia, das finanças pessoais à política monetária, podendo exercer efeito na expectativa dos investidores nos mercados financeiros como resposta a este fenómeno.

O *Federal Reserve*, o banco central dos Estados Unidos da América e mais conhecido como *Fed*, como principal interveniente na definição das políticas monetárias exerce influência na determinação das taxas de juro aplicadas nos mercados económicos e financeiros. Este exercício pode levar os investidores a alterar a composição do seu capital investido e procurar alternativas de menor risco, impactando negativamente índices de ações como o *S&P 500*.

O *S&P 500*, abreviação de *Standard & Poor's 500*, é um dos índices de ações mais amplamente seguidos e respeitados do mercado financeiro dos Estados Unidos uma vez que inclui as 500 empresas mais cotadas de capital aberto dos Estados Unidos, abrangendo uma variedade de setores da economia. Deste modo, as expectativas dos investidores no *S&P 500* desempenham um papel significativo nos movimentos do mercado uma vez que a inflação poderá implicar a redução no poder de investimento.

O objetivo da presente dissertação é a análise e previsão do impacto da inflação séries temporais no índice *S&P 500* utilizando como indicadores da inflação o *Consumer Price Index*, *Gross Domestic Product* e o *Unemployment Rate* através de instrumentos econométricos financeiros, incluindo modelos que têm a capacidade de prever valores de séries temporais.

Deste modo, este trabalho dividiu-se em três partes: No primeiro capítulo pretende-se ilustrar o conceito de inflação e a importância do seu estudo na atualidade bem como a evolução do índice *S&P 500*. O segundo capítulo destina-se ao suporte teórico da dissertação e é dividido em três partes: numa primeira fase são apresentadas as séries temporais e conceitos econométricos como estacionaridade e testes de raiz unitária. Posteriormente, são apresentados os modelos que serão utilizados para prever as variáveis em estudo. Por último, introduz-se o conceito de previsão, que através da aplicação de métodos de medida do erro, será possível a comparação dos valores previstos pelos modelos escolhidos com os valores reais.

O terceiro capítulo é reservado exclusivamente ao estudo empírico das variáveis. Isto é, serão aplicados os conhecimentos teóricos presentes no capítulo anterior com o auxílio da linguagem de programação *Jupyter Notebook*.

Por fim, é realizada uma síntese do estudo, onde será apresentada a comparação entre as previsões e os valores reais, de modo a alcançar o melhor modelo de previsão para as variáveis e posteriormente verificar a sua correlação.

Revisão da Literatura

1.1. Enquadramento teórico

1.1.1. O conceito de inflação

A inflação é o aumento de longo prazo nos preços de serviços e bens devido à desvalorização da moeda. Se os rendimentos não aumentarem à medida que a inflação aumenta, o poder de investimento será efetivamente reduzido, o que por sua vez conduzirá a uma economia estagnada ou lenta. Frequentemente expressa em percentagem, a inflação indica uma diminuição no poder de compra da moeda de uma nação.

Embora a inflação seja mais conhecida como um fenómeno com impacto negativo na perspetiva do consumidor local, por outro lado é bem recebida aos olhos de investidores em mercados cambiais e mercados financeiros que incluam Títulos do Governo associados à Inflação, chamadas de TIPS, como por exemplo.

A inflação é medida através da taxa de inflação designada como a variação percentual no preço dos produtos e serviços de um ano para o outro (ano após ano).

Duas das formas mais comuns de medir a inflação são:

- *Consumer Price Index (CPI)*, calculado pelo *Bureau of Labor Statistics* e o *Gross Domestic Product (GDP)* do *Bureau of Economic Analysis*. O CPI mede a mudança nos preços pagos pelos consumidores dos EUA ao longo do tempo e é a medida mais utilizada de avaliar a inflação^[3]. A taxa de inflação homóloga é calculada subtraindo o valor do CPI no início do ano do valor no final do ano. O resultado é então dividido pelo valor do CPI no início do ano e depois multiplicado por 100 para obter a percentagem da taxa de inflação.
- *Gross Domestic Product (GDP)*, é o valor monetário ou de mercado total de todos os bens e serviços acabados produzidos dentro das fronteiras de um país num período de tempo específico. Como medida ampla da produção interna global, funciona como um quadro de resultados abrangente da saúde económica de um determinado país.

Embora o GDP seja normalmente calculado numa base anual, por vezes também é calculado numa base trimestral. Nos EUA, o governo divulga uma estimativa anualizada do GDP para cada trimestre fiscal e também para o ano civil ^[4].

- *Unemployment Rate*, a inflação e o *Unemployment* normalmente têm uma correlação inversa, mas a relação é complexa. A inflação e o desemprego têm mantido historicamente uma relação inversa, representada pela curva de Phillips. Níveis baixos de desemprego correspondiam normalmente a uma inflação mais elevada, enquanto um desemprego elevado correspondia a uma inflação mais baixa e até a uma deflação.

1.1.2. A evolução da inflação nos Estados Unidos da América

Embora os Estados Unidos tenham registado uma taxa de inflação relativamente baixa e estável desde a década de 1980, a inflação atingiu máximos recordes em 2021 e 2022, na sequência da pandemia. A taxa de inflação anual foi de 7,0% em dezembro de 2021 e de 6,5% no final de 2022.

Observar a taxa de inflação do CPI do final de dezembro vincula-a a um momento específico, tornando mais fácil fazer comparações entre as taxas de inflação ano após ano (YoY). O gráfico abaixo apresenta a evolução da taxa de inflação anual nos EUA de 1929 a 2020.

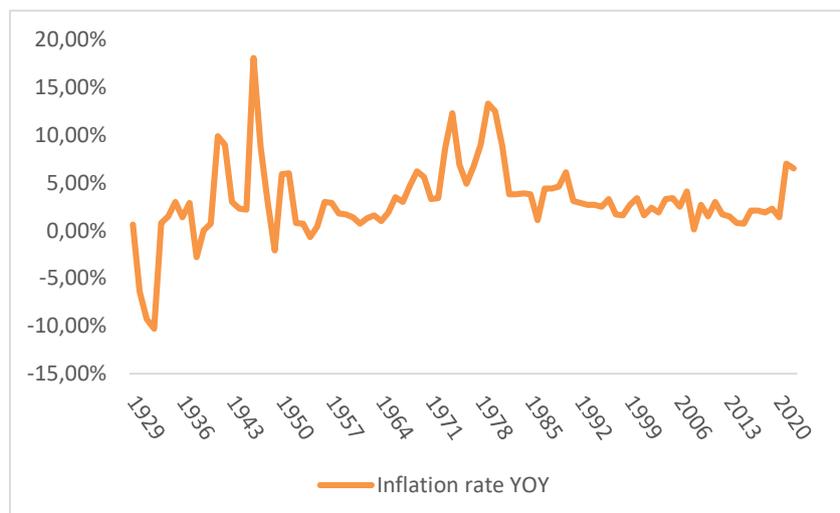


Gráfico 1 - Taxa de Inflação entre 1929 e 2020 nos EUA

Como é possível observar, existem eventos importantes que determinaram a subida e a descida acelerada da taxa de inflação nomeadamente:

- 1932 (-10,30%): Aumento dos impostos Hoover;
- 1946 (18,10%): Custos com a Guerra Fria;
- 1980 (12,80%): Recessão;
- 2021 (7,00%): Surto COVID-19;
- 2022 (6,50%): Guerra entre Ucrânia e Rússia.

1.1.3. O índice S&P 500

O S&P 500 é um índice de ações que acompanha os preços das ações de 500 das maiores empresas públicas dos Estados Unidos. Formalmente conhecido como Índice Composto de Preços de Ações da Standard & Poor's 500 e usualmente referido como S&P 500, é uma das principais ferramentas usadas para acompanhar o desempenho das ações dos EUA.

O S&P 500 acompanha os preços das ações de grande capitalização dos EUA, ou ações de empresas cujo total de ações em circulação valem mais de 10 bilhões US dólares. É por isso que o S&P 500 é frequentemente tratado como um proxy para descrever a saúde geral do mercado de ações ou mesmo da economia dos EUA.

As 500 maiores empresas públicas dos EUA por capitalização de mercado estão representadas no S&P 500, independentemente do setor em que atuam.

O S&P 500 é ponderado pela capitalização de mercado, pelo que a participação de cada empresa no índice global baseia-se no valor total de mercado de todas as suas ações em circulação. Os constituintes com maiores capitalizações de mercado têm uma ponderação percentual mais elevada no índice, enquanto as menores capitalizações de mercado têm ponderações mais baixas.

As seguintes empresas compõem os 10 principais constituintes do S&P 500 por peso do índice, de acordo com os índices S&P Dow Jones.

Company	Stock Ticker	Sector
Apple	AAPL	Information Technology
Microsoft	MSFT	Information Technology
Amazon.com	AMZN	Consumer Discretionary
Nvidia	NVDA	Information Technology
Alphabet (class A)	GOOGL	Communication Services
Tesla	TSLA	Consumer Discretionary

Meta Platforms (class A)	META	Communications Services
Alphabet (class C)	GOOGL	Communications Services
Berkshire Hathaway (class B)	BRK.B	Financials
Unitedhealth Group	UNH	Health Care

Tabela 1 - 10 empresa mais ponderadas no S&P 500

O gráfico abaixo apresenta a evolução dos retornos anuais em percentagem do S&P 500 nos EUA de 1957 a 2022.

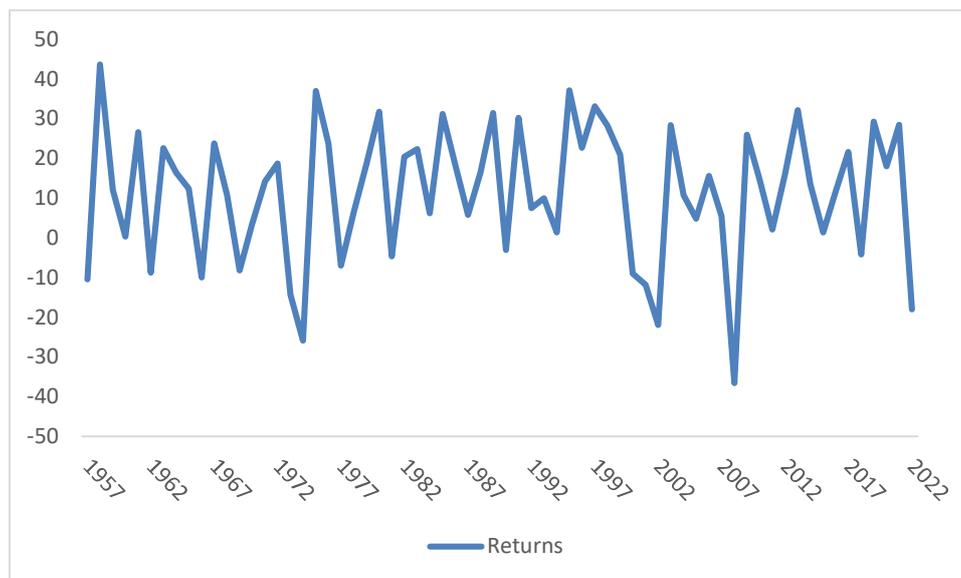


Gráfico 2 - Stock returns do S&P 500 entre 1957 e 2022

Como é possível observar, existem eventos importantes que estimularam a subida e/ou a descida dos retornos do S&P500 nomeadamente:

- Durante a primeira década após a sua introdução em 1957, e refletindo a expansão económica nos EUA após a Segunda Guerra Mundial, o valor do índice subiu para pouco mais de 800;
- De 1969 a 1981, o índice diminuiu gradualmente até cair abaixo de 360, como sinal de inflação elevada. Durante a crise financeira de 2008 e a Grande Recessão, o S&P 500 caiu 46,13% entre outubro de 2007 e Março de 2009. Em março de 2013, o S&P recuperou da crise e continuou a sua corrida de alta de 10 anos, de 2009 a 2019, para subir mais de 250%;
- A pandemia de COVID-19 em 2020 e a recessão subsequente fizeram com que o S&P 500 despencasse quase 20%;

- O S&P 500 recuperou durante o segundo semestre de 2020, atingindo vários máximos históricos em 2021, mas caiu mais de 1.100 pontos em 2022 antes de recuperar em junho;

A inflação é um dos principais problemas para um investidor que espera recriar regularmente aquele retorno médio de 11,88%. Ajustado pela inflação, o retorno médio anual histórico é de apenas cerca de 8,5%.

1.1.4. O impacto da inflação no S&P 500

À primeira vista, pode não ser óbvio se o aumento dos preços é mau para as ações. Embora a inflação elevada possa ter consequências negativas graves para a economia em geral, não é um desastre para os investidores.

Uma inflação mais elevada por si só não é necessariamente má para os preços das ações. O aumento dos preços aumenta os lucros das empresas, especialmente se as empresas puderem transferir custos mais elevados dos fatores de produção para os seus clientes através de aumentos de preços. Taxas de juros mais altas são uma história totalmente diferente para as ações. Quando a inflação foge ao controlo, a solução são taxas de juro mais altas, e o aumento das taxas torna o crédito mais caro para as empresas e os consumidores, desencorajando-os de gastar e investir. Isto pode reduzir os lucros e prejudicar as receitas, prejudicando os preços das ações.

Jamie Cox, sócio-gerente do Harris Financial Group, diz que a tendência negativa recente do mercado de ações tem mais a ver com as taxas de juros do que com a inflação. “Os mercados tendem a preocupar-se mais com a solução para a inflação – aumentos das taxas de juro – do que com a própria inflação”, diz Cox. “Os mercados descontam os lucros e fazem ajustes nos múltiplos com base no nível e na taxa de variação das taxas de juros, de modo que a cura aparece muito rapidamente nos mercados.”

Para atingir o objetivo, de fazer com que a inflação volte a cerca de 2% – exatamente onde o mercado deseja, são necessários aumentos de taxas, o que o mercado acionista definitivamente não gosta. A inflação elevada tende a gerar taxas de juros mais altas, o que tem um impacto negativo nas ações do S&P 500.

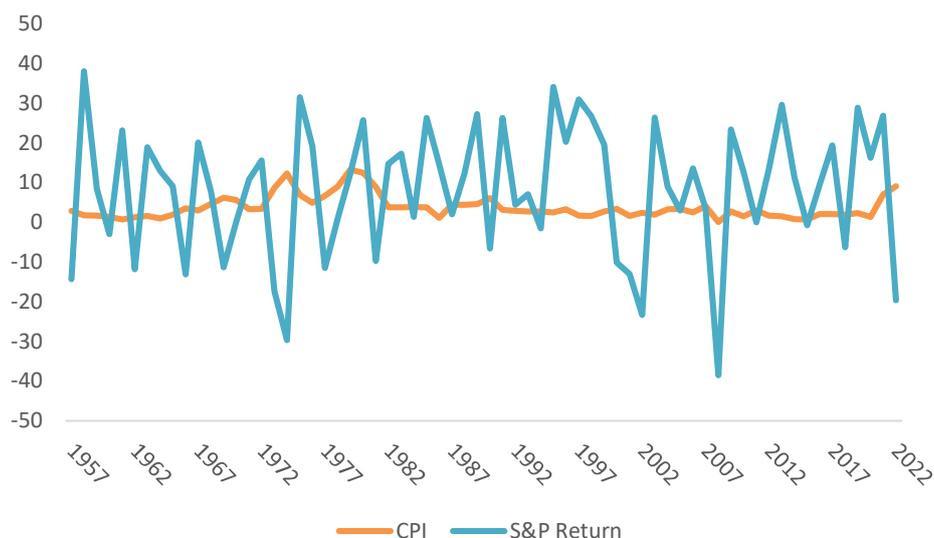


Gráfico 3 - Relação entre o CPI e o S&P 500 entre 1957 e 2022

Num mundo ideal, o mercado de ações gosta de ver um crescimento sustentado dos preços de cerca de 1% a 3% ao ano, o que é considerado uma inflação baixa a moderada. Este ambiente “saudável” significa que o valor do dólar permanece mais estável, a procura de bens e serviços mantém-se estável e os preços são previsíveis.

Mas quando a inflação anual sobe acima deste nível, o mercado de ações é dominado pela incerteza, pela volatilidade e pela desaceleração dos gastos dos consumidores. Isto, por sua vez, leva a um menor crescimento económico, o que é uma má notícia para os investidores e causa preocupações de avaliação que podem levar a um fraco desempenho.

Metodologia

Antes de se proceder à análise e interpretação das abordagens práticas presentes no próximo capítulo, é importante falar de alguns conceitos teóricos essenciais com o objetivo de entender os tópicos em estudo.

De notar que a elaboração do presente capítulo será assistida nomeadamente pelos livros (Brooks, 2019) e (Mills, 2019), uma vez que ambos apresentam uma excelente base teórica para o desenvolvimento deste capítulo.

2.1. Séries Temporais

Uma série temporal é um conjunto de observações que aparecem numa forma sequencial, normalmente com um intervalo temporal fixo entre as observações. Esses intervalos temporais podem ser diários, semanais, mensais, etc., sendo que a ordem das observações é crucial no estudo de séries temporais.

Em termos matemáticos, uma série temporal pode ser definida como discreta

$$Y_t, t = 1, 2, \dots, T$$

ou pode ser definida como contínua

$$Y_t, t \in T \subseteq \mathbb{R} \quad (2)$$

As séries temporais que apresentam uma única observação para um determinado momento temporal caracterizam-se como séries temporais univariadas, em contrapartida as que apresentam duas ou mais observações por instante no tempo são chamadas de séries temporais multivariadas. Esta dissertação irá abordar o estudo de séries temporais univariadas.

Cada série temporal contém 4 componentes/movimentos principais: movimentos ou variações por estações (sazonalidade), movimentos ou variações cíclicas, movimentos de tendência e movimentos irregulares ou aleatórios.

Os movimentos de tendência são movimentos que se referem à direção global na qual a série temporal se desenvolve, num determinado período temporal. Este tipo de movimento pode ser determinado através de uma regressão linear. Os movimentos ou variações cíclicas são variações de longo prazo periódicos ou não periódicos em torno da curva de tendências. Um bom exemplo de variações cíclicas é o ciclo dos negócios, pois apresentam períodos de recessão, depressão, prosperidade e recuperação. Os movimentos ou variações por estações referem-se a movimentos iguais ou similares que a série apresenta durante os mesmos períodos temporais consecutivos. Um exemplo de sazonalidade é o crescimento nas vendas de protetor solar durante o verão. Os movimentos irregulares ou aleatórios referem-se aos desvios esporádicos das séries temporais, instigados por fatores externos como epidemias ou guerras. Estes desvios provocam variações durante um curto intervalo, no entanto estas variações poderão ser tão intensas ao ponto de criar movimentos cíclicos ou de outra natureza.

2.2. Descrição das séries temporais

Para examinar melhor as series temporais existem estatísticas que ajudam o analista a ter uma melhor interpretação das características que a série apresenta.

Um processo estocástico é definido por qualquer família de variáveis aleatórias, $Y_t, t \in T$ na qual T corresponde a um período predefinido.

Seja $Y(t)$ um processo estocástico com variância finita para $\forall t \in T$, as suas principais estatísticas definem-se em:

- Média ou Valor Esperado, $\mu(t)$, em que: $\mu(t) = E[Y(t)] = E[Y(t)]$
- Variância, $\sigma^2(t)$, em que: $\sigma^2(t) = Var[Y(t)] = E[(Y(t) - \mu(t))^2]$;
- Covariância, $\gamma(t_1, t_2)$, em que: $\gamma(t_1, t_2) = Cov[Y(t_1), Y(t_2)] = E\{[Y(t_1) - \mu(t_1)][Y(t_2) - \mu(t_2)]\} = E[Y(t_1)Y(t_2)] - \mu(t_1)\mu(t_2)$;
- Correlação, $\rho(t_1, t_2)$, em que: $\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sigma(t_1)\sigma(t_2)}$

2.2.1. Assimetria

A assimetria é uma das medidas frequentemente usadas na análise de dados financeiros. Em termos estatísticos, a assimetria mostra o grau de afastamento que uma determinada distribuição apresenta do seu eixo de simetria. Para calcular o coeficiente de simetria, S_k , utiliza-se a fórmula:

$$s_k = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^3}{\left[\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{\frac{3}{2}}} \quad (3)$$

onde n é o número de observações da série temporal y_i é o valor observado da mesma no momento i e \bar{y} é a média amostral.

Depois de calculado o valor do coeficiente de simetria, existem três conclusões possíveis relativamente a esse valor. Se o coeficiente de simetria for igual a zero, significa que a distribuição é simétrica e a Média = Mediana = Moda. Se o coeficiente for positivo, a distribuição é assimétrica positiva (à direita), ou seja, Média > Mediana > Moda. Por último, no caso em que o coeficiente é negativo, a distribuição é assimétrica negativa (à esquerda), o que significa que Média < Mediana < Moda.

2.3. Estacionaridade

A análise da estacionaridade de uma série temporal é um passo fulcral no estudo das mesmas, já que é uma condição essencial para a utilização de alguns modelos e para a previsão das séries temporais.

Existem dois tipos de estacionaridade: estacionaridade estrita (forte) ou estacionaridade fraca. Um processo estocástico diz-se estritamente estacionário se as suas propriedades não são afetadas por mudanças no tempo (Mills, 2019).

Quando se prova estacionaridade, geralmente refere-se a estacionaridade fraca, já que é difícil achar uma série temporal estritamente estacionária.

Um processo estocástico $Y(t), t \in T$ diz-se fracamente estacionário se, $\forall t \in T$, se tem:

- $E[Y(t)] = \mu(t)$
- $Var[Y(t)] = \sigma^2(t) < \infty$
- $Cov[Y(t_1), Y(t_2)] = \gamma(t_1, t_2) = \gamma(|t_1 - t_2|), t_1, t_2 \in T$

Então, uma série temporal é fracamente estacionária se a sua média for constante ao longo do tempo, se a sua variância for constante e finita ao longo do tempo e se a sua covariância entre os valores desfasados desta série depender apenas do valor da distância temporal entre eles. Uma série temporal que não verifique estes critérios diz-se não-estacionária. Para além disso, uma série estritamente estacionária significa que ela é automaticamente fracamente estacionária, no entanto o contrário pode não se verificar (Mills, 2019).

Quando não se verifica estacionaridade numa série temporal existem certos procedimentos que poderão ajudar a estabilizar as mesmas e torná-las estacionárias:

2.3.1. Diferenças Sucessivas

$$\text{Primeira diferença: } \Delta Y_t = Y_t - Y_{t-1}$$

$$\text{Segunda diferença: } \Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

2.3.2. Diferenças Logaritmizadas

$$\Delta \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$$

Se uma série não-estacionária for diferenciada d vezes antes de se tornar estacionária, diz-se que essa série é integrada de ordem d , denotada por $Y_t \sim I(d)$. Uma série estacionária que não necessite de ser diferenciada diz-se que é integrada de ordem 0, $I(0)$.

2.3.3. Ruído Branco

Um processo $\{\varepsilon_t\}$ define-se como ruído branco se:

- $E(\varepsilon_t) = 0$
- $Var(\varepsilon_t) = \sigma^2$
- $E(\varepsilon_j, \varepsilon_k) = 0, j \neq k$

Então, um processo define-se como ruído branco se tiver média nula, variância constante e finita e não apresentar correlação serial. Um processo de ruído branco representa-se por $\varepsilon_t \sim WN(0, \sigma^2)$. Para além disso, os erros de modelos devem ser ruídos brancos, já que o ideal é não existir correlação entre eles. Se os erros forem correlacionados, a modelização e previsão de valores podem não ser confiáveis.

As séries temporais podem apresentar dois tipos de tendência: tendência determinística, em que existem variações aleatórias em torno da curva de tendência, não impactando o comportamento do processo no longo prazo, ou tendência estocástica, em que a componente aleatória ε_t influencia o comportamento da série no longo prazo.

Considera-se o seguinte modelo

$$Y_t = \rho Y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t$$

e os seguintes cenários:

- Se $\rho = 0$, significa que temos um processo estocástico em tendência puro (desenvolve-se em torno de uma tendência linear – média não constante), ou seja, existe uma tendência determinística. Neste caso, a série é não-estacionária;
- Se $\rho = 1$ e $\beta_1 = 0$, significa que temos um processo estocástico em tendência puro (desenvolve-se em torno de uma tendência linear – média não constante), ou seja, existe uma tendência determinística. Neste caso, a série é não-estacionária;
- Se $\rho = 1$ e $\beta_0, \beta_1 = 0$ então temos um processo estocástico em diferenças puro, também denominado por passeio aleatório com drift, logo existe uma tendência estocástica. Neste caso, a série é não-estacionária;
- Se $\rho = 1$ e $\beta_0, \beta_1 \neq 0$, significa que o processo estocástico é um passeio aleatório com drift e tendência determinística. A série é não-estacionária;
- Se $0 < \rho < 1$, significa que existe correlação serial em Y , mas não existe tendência estocástica. A série é estacionária.

Para sintetizar, o estudo da não-estacionaridade de uma série é analisar se $\rho = 1$. Caso se verifique, a série apresenta uma raiz unitária e é não-estacionária.

2.3.4. Testes de Raiz Unitária

Geralmente as séries temporais são não-estacionárias, no entanto esta característica tem de ser provada. Para este estudo foram usados testes para detetar se a série temporal é estacionária ou não. Se a série não for estacionária diz-se que tem uma raiz unitária.

É então testada a hipótese nula de que a série é não-estacionária ($H_0: \rho = 1$, a série é não-estacionária) versus a hipótese de que a série é estacionária ($H_0: \rho \neq 1$, a série é estacionária).

Existem vários testes de raiz unitária, no entanto os mais frequentes são: o teste DF (Dickey-Fuller), o teste ADF (Augmented Dickey-Fuller), o teste PP (Phillips-Perron) e o teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin).

O teste de Dickey-Fuller é utilizado quando se assume que os erros (ε_t) são ruído branco, isto é, quando os erros são independentes e não apresentam autocorrelação. Ao utilizar o modelo previamente exemplificado, presente na equação (2.6), subtrai-se Y_{t-1} em ambos os lados da equação:

$$\Delta Y_t = (\rho - 1)Y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t$$

Para o teste DF, as hipóteses testadas são ($H_0: \rho = 1$, existência de pelo menos 1 raiz unitária) versus ($H_0: \rho < 1$, não existe raiz unitária).

A condição de não correlação serial entre os erros nem sempre é verificada, portanto é necessário um teste que faça a correção paramétrica. Este teste é chamado de teste de Dickey-Fuller Aumentado (ADF), onde a regressão é aumentada em ΔY_{t-i} termos, ao ponto de garantir que os erros não apresentem autocorrelação. O modelo de exemplo modifica-se novamente:

$$\Delta Y_t = (\rho - 1)Y_{t-1} + \sum_{i=1}^p \beta_i \Delta Y_{t-i} + \varepsilon_t$$

As hipóteses referentes ao teste ADF são idênticas às do teste DF. O teste de Phillips-Perron é semelhante aos dois testes anteriores, no entanto contém uma correção automática, permitindo aos erros serem autocorrelacionados. As hipóteses referentes ao teste PP são idênticas às do teste ADF.

Por último, o teste de KPSS, chamado de teste de estacionaridade, foi desenvolvido com o objetivo de ser um contraposto aos testes de raiz unitária usuais. Este teste considera as seguintes hipóteses ($H_0: \rho = 1$, a série é não-estacionária) versus a hipótese de que a série é estacionária ($H_0: \rho < 1$, a série é estacionária).

Seja $s_t = \sum_{i=1}^t \varepsilon_i$ e $\hat{\sigma}^2$ um estimador para a variância dos erros, a estatística de teste é dada por:

$$LM = \sum_{t=1}^T \frac{s_t^2}{n^2 \hat{\sigma}^2}$$

Neste teste rejeita-se H_0 se a estatística de teste $LM_{KPSS} >$ valores críticos usuais. O teste KPSS é geralmente usado como complemento dos testes apresentados anteriormente, já que apresenta uma diferença significativa nas hipóteses testadas. E também frequentemente utilizado como critério de desempate na discordância entre os testes ADF e PP.

2.4. Modelos Univariados

Os modelos univariados de séries temporais utilizam unicamente os valores passados das mesmas para modelizar e prever observações futuras. Como foi referido no capítulo anterior, um desses modelos é o modelo ARMA, usado para modelizar séries estacionárias. Para séries não-estacionárias, aplica-se o

modelo ARIMA, onde o I é referente ao processo de integração de uma série não estacionária, com o auxílio de diferenças sucessivas ou diferenças logaritmizadas.

Os modelos da família ARMA/ARIMA assumem variância dos erros constante ao longo do tempo (homocedasticidade). Em contrapartida, os modelos da família ARCH/ GARCH conseguem captar a heterocedasticidade condicional nos erros do modelo, captando a volatilidade das séries temporais. Como já foi referido, existe uma extensão destes modelos, o modelo ARIMA-GARCH, modelo este que incorpora o modelo da família ARMA/ARIMA com o modelo GARCH, conseguindo assim captar a volatilidade das séries temporais (heterocedasticidade).

O ARIMA-GARCH é utilizado para modelizar a média com o modelo ARIMA e a variância com o modelo GARCH. No entanto, as previsões obtidas por ambas as equações irão ser somadas para criar a previsão final deste modelo, sendo este um procedimento não muito frequente no estudo de séries temporais financeiras com modelos ARIMA-GARCH.

2.4.1. Modelos Autorregressivos - $AR(p)$

Os modelos autorregressivos são modelos nos quais o valor da variável dependente Y_t depende apenas do valor dessa mesma variável em períodos passados, somado com um termo de erro. Com isto, o modelo autorregressivo de ordem $p - AR(p)$ é escrito da seguinte forma:

$$Y_t = a_0 + \sum_{i=1}^p \beta_i \Delta Y_{t-i} + \varepsilon_t \sim WN(0, \sigma^2)$$

Onde a_0, a_1, \dots, a_p são os coeficientes do modelo, p é a ordem de desfasamento, ε_t é um processo de ruído branco.

Seja L o operador de desfasamento, representado por $L(Y_t) = Y_{t-1}$. O modelo anterior pode ser reescrito em:

$$\phi L(Y_t) = a_0 + \varepsilon_t$$

onde

$$\phi L = 1 - \sum_{i=1}^p a_i L^i$$

Um processo autorregressivo de ordem p estacionário caso todas as raízes $z = \frac{1}{\alpha_i}$ do polinómio

$$\phi(z) = 1 - \sum_{i=1}^p a_i z^i$$

estão situadas no exterior do círculo unitário, isto é, se $|\alpha_i| < 1$.

2.4.2. Modelos de Média Móvel - MA(q)

Os modelos de média móvel são feitos através de uma combinação linear de processos de ruído branco, de modo que a variável dependente Y_t seja influenciada pelos valores atuais e passados de um termo de distribuição de ruído branco.

Seja ε_t um processo de ruído branco. O processo MA(q) pode ser representado por:

$$Y_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

Ao utilizar o operador de desfasamento, o processo modifica-se novamente:

$$Y_t = \sum_{i=1}^q \theta_i L^i \varepsilon_t = \theta_i(L) \varepsilon_t$$

onde $\theta_i(L) = 1 + \sum_{i=1}^q \theta_i L^i$ e θ_i são constantes reais. Um processo MA(q) é fracamente estacionário para a totalidade dos valores de θ_q .

2.4.3. Modelos Autorregressivos de Médias Móveis – ARMA (p, q)

Ao combinar o modelo autorregressivo de ordem p , AR(p), com o modelo de média móvel de ordem q , MA(q), obtém-se o modelo autorregressivo de médias móveis, ARMA (p, q).

Este modelo afirma que o valor atual de uma variável Y_t depende linearmente dos seus valores passados (AR), somado a uma combinação de valores presentes e históricos de um termo de erro de ruído branco (MA). Sendo assim, o modelo ARMA(p, q) define-se pela seguinte equação:

$$Y_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{k=1}^q \theta_k \varepsilon_{t-k} + \varepsilon_t$$

Este modelo é muito eficaz no que toca à previsão de valores de séries temporais, no entanto tem certas limitações. Uma dessas limitações é o assumir que todas as séries temporais são estacionárias, o que não é verdadeiro, já que a maioria das séries temporais são efetivamente não-estacionárias.

2.4.4. Modelos Autorregressivos Integrados de Médias Móveis - ARIMA(p, d, q)

Os modelos do grupo ARMA assumem que todas as séries temporais são estacionárias, algo que pode não ser verificado. Foi então definido o modelo $ARIMA(p, d, q)$, onde o d representa o número de vezes que a série temporal teve de ser diferenciada até atingir estacionaridade, o p representa a ordem da AR e o q representa a ordem da componente MA.

É também importante referir que os modelos univariados previamente apresentados estão inseridos dentro do modelo $ARIMA(p, d, q)$, isto é:

- $ARIMA(p, 0, q) = ARMA(p, q)$
- $ARIMA(p, 0, 0) = AR(p)$
- $ARIMA(0, 0, q) = MA(q)$

2.4.5. A Metodologia de Box-Jenkins

A metodologia de Box-Jenkins foi desenvolvida com o objetivo de estimar modelos do tipo ARMA/ARIMA de uma forma sistemática. Esta metodologia segue três passos:

- Passo 1 – Identificação

Este passo consiste na determinação da ordem do modelo necessária para capturar as características dos dados. Para este passo são utilizadas as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) como auxílio gráfico e os critérios de informação, como o AIC (Akaike Information Criterion) e o BIC (Bayesian Information Criterion), com o objetivo de escolher o número de parâmetros que minimizam o valor dos critérios de informação. No entanto, a adição de um termo extra só diminui o valor do critério de informação se o decréscimo na soma dos resíduos ao quadrado for mais acentuado que o acréscimo do valor do termo de penalidade. As equações para o AIC e BIC são dadas por, respetivamente,

$$AIC = \ln \hat{\sigma}^2 + \frac{2k}{T}$$

$$BIC = \ln \hat{\sigma}^2 + \frac{k}{T} \ln(T)$$

onde $\hat{\sigma}^2$ é a variância residual, que é equivalente à divisão da soma dos resíduos ao quadrado pelo número de observações T , e $k = p + q + 1$, que é o número total de parâmetros estimados.

- Passo 2 - Estimação

Neste passo são estimados os parâmetros do modelo previamente escolhido. Esta estimação pode ser feita através do método dos mínimos quadrados ou de Máxima Verossimilhança. A escolha do tipo de método vai depender das características do modelo.

- Passo 3 - Diagnóstico

Neste passo deve-se verificar se o modelo estimado é adequado, sendo necessária uma análise aos resíduos.

Para que o modelo seja validado, ele tem que verificar os seguintes pressupostos:

- $E(\varepsilon_t) = 0$: erros têm média nula;
- $Var(\varepsilon_t) = \hat{\sigma}^2 < \infty$: variância dos erros constante e finita (homocedasticidade). Caso contrário, os erros do modelo são heterocedásticos;
- $Cov(\varepsilon_i, \varepsilon_t) = 0, i \neq j$: não existe qualquer correlação entre os erros ao longo do tempo, sendo eles linearmente independentes;
- $\varepsilon_t \sim N(0, \hat{\sigma}^2)$: os erros têm de ser normalmente distribuídos.

2.4.6. Modelos ARCH/GARCH

Como já foi referido anteriormente, os modelos univariados lineares apresentados até agora (ARMA/ARIMA) assumem que a variância dos erros é constante ao longo do tempo, algo pouco provável quando se trata de séries financeiras. Outros problemas que podem surgir com modelos desta família é a incapacidade de explicar certas características financeiras importantes, como a leptocurtose, os clusters de volatilidade e o efeito de alavanca.

Com o intuito de tentar solucionar alguns desses problemas, foi criado o modelo ARCH(q) (Autoregressive Conditional Heteroskedasticity), desenvolvido por Engle em 1982. Posteriormente, uma extensão do modelo ARCH foi definida, o chamado modelo GARCH(p, q) (Generalized Autoregressive Conditional Heteroskedasticity).

2.4.7. O modelo GARCH

Introduzido por Bollerslev em 1986, este modelo permite à variância condicional ser dependente dos seus próprios desfasamentos.

O modelo GARCH(p, q) define-se por duas equações: a equação da média

$$y_t = \phi_1 + \phi_2 y_{t-1} + \dots + \phi_k y_{t-k} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2)$$

e a equação da variância

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2$$

onde σ_t^2 é a variância condicionada, α_0 é a constante, ε_{t-q}^2 é a volatilidade do período anterior e σ_{t-p}^2 é a variância do período anterior. Para a equação da variância, os coeficientes devem ser não-negativos.

É também importante referir que um modelo do tipo GARCH(0, q) é equivalente a um modelo ARCH(q).

No que segue, e sem perda de generalidade, vai ser apresentado o modelo GARCH(1,1), considerado o modelo baseline para os retornos de séries financeiras:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

onde σ_t^2 é a variância condicional, α_0 é a constante, ε_{t-1}^2 é a volatilidade do período anterior e σ_{t-1}^2 é a variância do período anterior.

No modelo GARCH(1,1), a variância condicional altera-se, mas a variância não condicional dos erros (u_t) é constante e é dada por

$$var(u_t) = \frac{\alpha_0}{1 - (\alpha_1 + \beta)}$$

desde que $\alpha_1 + \beta < 1$. Caso $\alpha_1 + \beta \geq 1$, então existe não estacionaridade em variância. Por último, caso $\alpha_1 + \beta = 1$, então o modelo GARCH é integrado (também chamado de IGARCH), o que significa que um choque na variância condicional vai ser importante não só no presente, como também em futuras observações.

Como foi discutido anteriormente, modelos do grupo GARCH não são lineares, logo não é possível estimar utilizando o Método dos Mínimos Quadrados. Para estimar modelos do tipo GARCH é utilizada a técnica de Máxima Verosimilhança (MV). Este método consiste em criar uma função log-verosimilhança e procurar pelos valores dos parâmetros que maximizam a mesma. Esta técnica consegue estimar parâmetros para modelos não lineares e para modelos lineares.

No entanto, os modelos GARCH têm certas limitações: eles não conseguem captar os efeitos de alavancagem, não permitem feedback entre média e variância condicional e as condições não negativas poderão ser infringidas com a estimação do modelo.

Para combater estas limitações, foram desenvolvidas várias versões do modelo GARCH, como o IGARCH ou o EGARCH, com o propósito de ganhar habilidade em modelizar os diferentes efeitos e características das séries temporais financeiras. Estas extensões do modelo GARCH não serão abordadas na presente dissertação.

2.5. Previsão

A última fase deste processo é a previsão de valores a partir dos modelos escolhidos. A fase de previsão está dividida em duas partes: a parte da previsão de valores futuros de uma série temporal e da comparação destes valores previstos com os valores reais da série temporal.

Numa primeira parte, será avaliada a capacidade de ajustamento dos modelos selecionados para as séries originais e para os retornos. A segunda parte passa por uma previsão *in-sample*, ou seja, é uma previsão dentro do intervalo da amostra. Geralmente as previsões *in-sample* são boas, por essa mesma razão.

Numa última parte e se as previsões *in-sample* para as séries em estudo forem boas, vai-se avançar com as previsões *out-of-sample*, que são previsões fora do intervalo de observações.

2.5.1. Erro de Previsão

O erro de previsão de uma observação no momento de tempo t é dado pela diferença entre o valor atual observado e o valor previsto em t , isto é

$$e_t = Y_t - \hat{Y}_t$$

onde e_t é o erro de previsão observado em t .

Na avaliação do desempenho global de previsões, foram criadas algumas medidas de erro, nomeadamente o Erro Percentual Absoluto Médio (MAPE), o Erro Absoluto Médio (MAE) e a Raiz do Erro Quadrático Médio (RMSE). Estas medidas são dadas pelas seguintes fórmulas

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{Y_t}$$

$$MAE = \frac{\sum_{t=1}^n |e_t|}{n}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}}$$

Com isto, quanto menores forem os erros de previsão, melhores serão as previsões executadas pelo modelo escolhido. E também importante referir que as previsões ficam cada vez menos eficazes caso se aumente a janela de previsão.

CAPÍTULO 3

Análise Empírica

Neste capítulo serão aplicados os conhecimentos teóricos apresentados anteriormente, com o objetivo de encontrar o modelo econométrico que melhor se ajuste aos dados das séries temporais selecionadas e que minimize o erro de previsão. Neste sentido, serão analisadas as quatro séries temporais *S&P 500*, *Consumer Price Index*, *Gross Domestic Product* e *Unemployment*, ambas com valores registados entre janeiro de 2015 e dezembro de 2022 e frequência trimestral, num conjunto de 60 observações. Todos os valores referentes às variáveis foram retirados da base de dados Federal Reserve Bank of St. Louis. É importante referir que os outputs, incluindo os gráficos e tabelas, neste capítulo foram obtidos com o auxílio da linguagem de programação *Jupyter Notebook*.

3.1. Análise das Séries Temporais

Para perceber melhor o comportamento das séries, é importante analisar graficamente as mesmas.

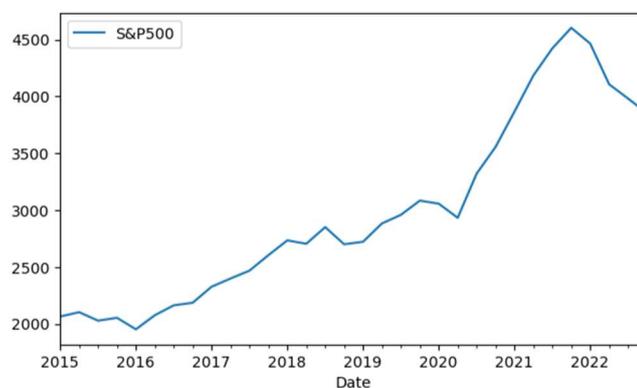


Figura 1 - Representação gráfica da série S&P 500

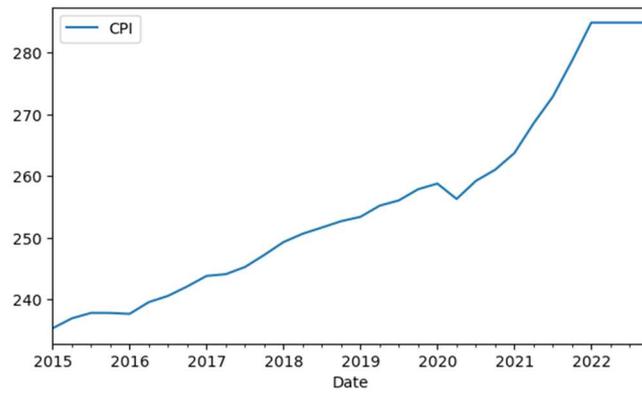


Figura 2 - Representação gráfica da série Consumer Price Index

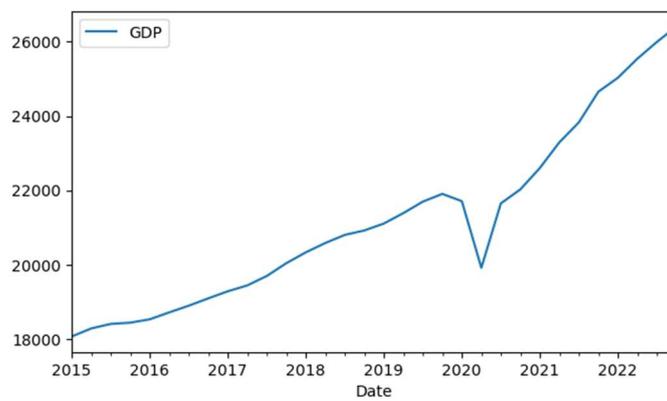


Figura 3 - Representação gráfica da série Gross Domestic Product

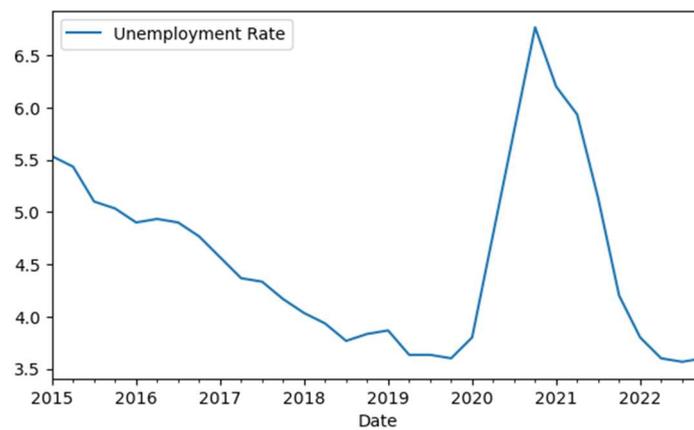


Figura 4-Representação gráfica da série Unemployment Rate

Ao analisar a figura 1 verifica-se que a série S&P 500 apresenta um padrão não linear, com deslocações irregulares e com média e variância não constantes, um forte indício de que a série é não-estacionária. As eleições presidenciais de 2016 nos Estados Unidos foram um dos eventos mais significativos do ano. A vitória de Donald Trump nas eleições em novembro de 2016 teve um impacto inicial positivo nos mercados, impulsionando as ações. Em 2020, o S&P 500 teve um declínio acentuado no início do ano em resposta à pandemia, o índice experimentou uma forte recuperação no segundo semestre de 2020, impulsionada por medidas de estímulo e a perspectiva de uma vacina para a COVID-19.

A série Consumer Price Index na figura 2 transmite um padrão linear, com apenas uma descida no período da pandemia de COVID-19 que teve um impacto significativo nos preços em 2020. No início da pandemia, houve uma queda nos preços de combustíveis, passagens aéreas e estadias devido à redução das viagens e da procura por combustíveis. Ao mesmo tempo, houve aumento nos preços de alimentos e produtos de cuidados pessoais devido a um aumento na procura.

Esta série apresenta uma tendência crescente como seria de esperar com a evolução da economia.

Na figura 3, a série *Gross Domestic Product* parece seguir o mesmo padrão da *Consumer Index Price* com um crescimento linear e com tendência crescente. Apenas o pico negativo precisamente em 2020. O consumo pessoal, que é um dos principais componentes do GDP, diminuiu significativamente devido à incerteza económica e às restrições à mobilidade. É importante observar que, em resposta à pandemia, as autoridades dos Estados Unidos adotaram uma série de políticas para mitigar os impactos económicos e promover a recuperação. A situação económica continuou a evoluir após 2020, com vários fatores, incluindo o progresso na vacinação e as políticas económicas, afetando a trajetória da economia dos EUA nos anos seguintes.

O *Unemployment Rate* na figura 4 parece também seguir o comportamento das duas variáveis anteriores, mas com o sinal contrário uma vez que se trata de uma variável que se pretende diminuir. Neste caso, não existe linearidade nem uma tendência a seguir por se registar o maior valor neste intervalo: 6,77% aproximadamente. É importante mencionar que a pandemia de COVID-19 atingiu os Estados Unidos no início do ano, levando a *lockdowns*, encerramento de empresas e restrições comerciais para conter a propagação do vírus. Como resultado das medidas de contenção da pandemia, houve um aumento brutal e significativo no desemprego dos Estados Unidos.

Completa-se a análise visual das séries temporais com as medidas de estatística descritiva que se fazem corresponder a cada uma das variáveis.

	SP500	GDP	Unemp	CPI
count	32.000000	32.000000	32.000000	32.000000
mean	2980.581314	21193.879219	4.546875	255.449375
std	816.694649	2417.446479	0.865990	15.520269
min	1951.224918	18063.529000	3.566667	235.355000
25%	2290.665570	19232.407750	3.800000	243.412667
50%	2791.539075	20858.298500	4.350000	253.062500
75%	3628.869545	22168.422750	5.050000	261.717000
max	4602.108906	26408.405000	6.766667	284.893667

Tabela 2 - Representação da Estatística Descritiva das séries em estudo

3.1.1. Estacionaridade

Como foi referido anteriormente, o estudo da estacionaridade é crucial para a análise de séries. Para o estudo da estacionaridade vão ser utilizados 2 tipos de testes, para cada uma das séries: o teste ADF e o teste KPSS. Os resultados obtidos para ambas as séries estão presentes nas tabelas seguintes:

Testes	S&P 500	CPI	GDP	Unemployment Rate
ADF	0.998	0.999	0.992	0.021
KPSS	0.010	0.010	0.010	0.100

Tabela 3 - Testes de Estacionaridade

Através da tabela 2, tanto no teste ADF como no KPSS, para as séries *S&P 500*, *CPI* e *GDP*, não se rejeita a hipótese nula uma vez que o p-value é maior que 0.05 (ADF) e se rejeita a hipótese nula uma vez que o p-value é menor que 0.05 (KPSS) e por isso as séries são não-estacionárias.

Já a série *Unemployment Rate* apresenta um p-value 0.021 (ADF) e 0.100 (KPSS), o que indica que a mesma é estacionária segundo o teste ADF e estacionária para o KPSS.

Assumindo que todas as variáveis são não-estacionárias, é necessário proceder-se à aplicação da diferença logarítmica para as séries *S&P 500*, *CPI* e *GDP*. Já a série *Unemployment Rate* foi aplicada a diferença normalizada. Deste modo, foi possível alcançar os valores da tabela abaixo.

Testes	S&P 500	CPI	GDP	Unemployment Rate
ADF	0.014	0.010	0.000	0.013
KPSS	0.100	0.100	0.100	0.100
PP	-1.411	-0.996	N/A	-1.440
d	1	1	1	1

Tabela 4 - Valor dos p-values nos testes de estacionaridade

Como é possível verificar, todas as variáveis à exceção do S&P 500 ficaram estacionárias após a primeira diferenciação. Contudo, para as variáveis *S&P 500*, *Consumer Price Index* e *Unemployment Rate* foi necessário utilizar o teste do *Phillips-Perron* para desempatar levando a *critical values* de -4.28, -3.56 e -3.22 para níveis de significância de 1%, 5% e 10%, respetivamente, já que os valores da estatística de teste foram inferiores aos dos valores críticos. Não se rejeita a hipótese nula da série ser estacionária, garantindo a estacionaridade da série.

Assim, temos que as variáveis são todas integradas de ordem $I(1)$, podendo-se proceder à estimação do modelo que se adapta melhor às características das nossas variáveis e capaz de realizar uma previsão com o menor erro.

3.2. Estimação do Modelo

Para um modelo ser capaz de aprender os padrões registados pelas variáveis que vai modelar e prever com o melhor desempenho, é necessário que a informação das séries temporais seja dividida em dois grupos: um de treino, onde o modelo se vai basear; um de teste, para posterior comparação com os resultados gerados pela aplicação do mesmo. Deste modo, decidiu-se dividir as séries em 16 observações cada um.

3.2.1. Estimação dos Modelos Univariados

Garantida a estacionaridade das séries, agora é importante escolher o modelo para continuar o estudo empírico. Para isso, será necessário olhar para a representação gráfica da função de autocorrelação (ACF) das séries para tentar perceber a ordem dos modelos univariados.

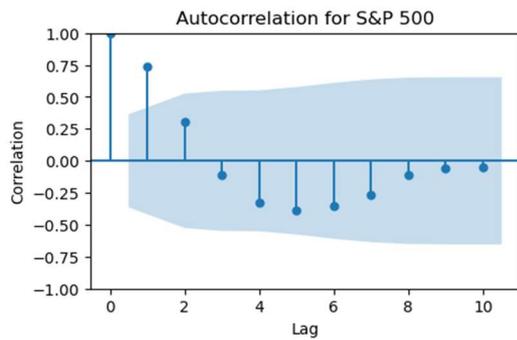


Figura 5 - Autocorrelação para S&P 500

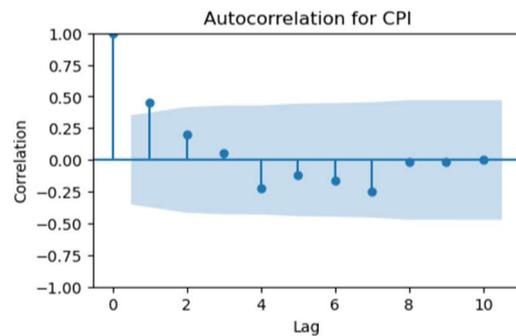


Figura 6 - Autocorrelação para o CPI

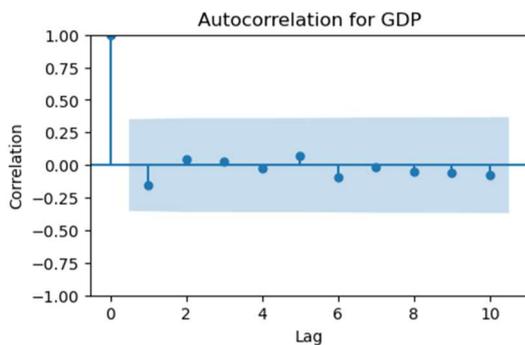


Figura 7 - Autocorrelação para o GDP

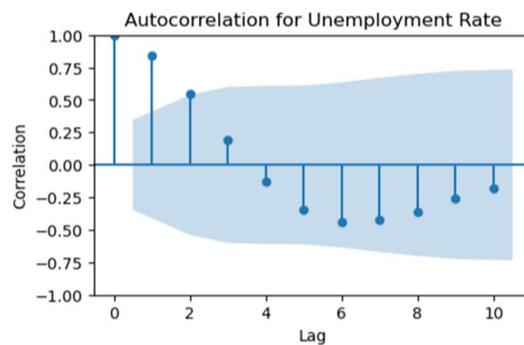


Figura 8 - Autocorrelação para o Unemployment Rate

Ao analisar as Figuras 5, 6, 7 e 8 percebe-se que só o primeiro *lag* é significativo.

Como foi referido em capítulos anteriores, analisar apenas estas funções não é suficiente para descobrir o modelo ótimo para o conjunto de dados. Para isso, foi utilizada uma função em código chamada *auto.arima*, que retorna o modelo ARIMA que minimiza os critérios de informação AIC e BIC, de modo a facilitar e a melhorar a escolha do modelo ideal.

Para cada uma das variáveis em estudo, foi possível obter os seguintes modelos ARIMA:

Variáveis	S&P 500	CPI	GDP	Unemployment Rate
Model	ARIMA(1,1,0)	ARIMA(0,1,1)	ARIMA(1,1,3)	ARIMA(1,1,0)

Tabela 5 - Modelos ARIMA obtidos através do *auto.arima*

Apresentando os seguintes valores previstos, graficamente:

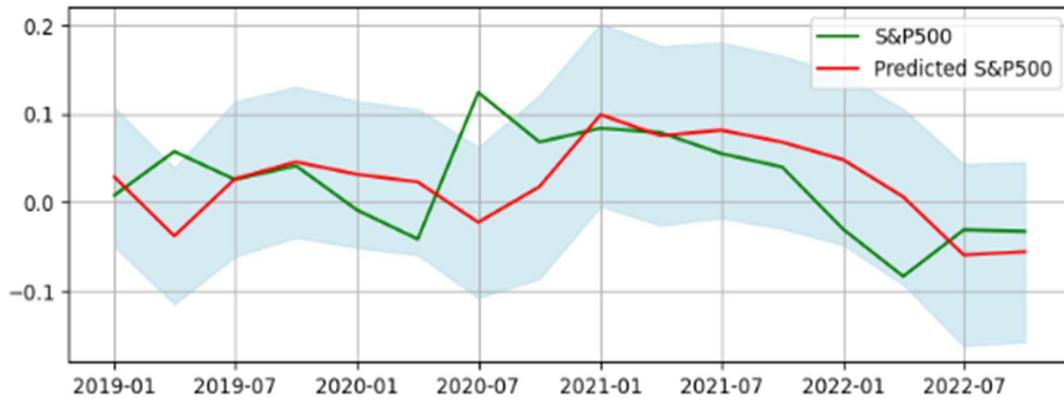


Figura 9 - Previsão segundo o ARIMA(1,1,0) para o S&P 500

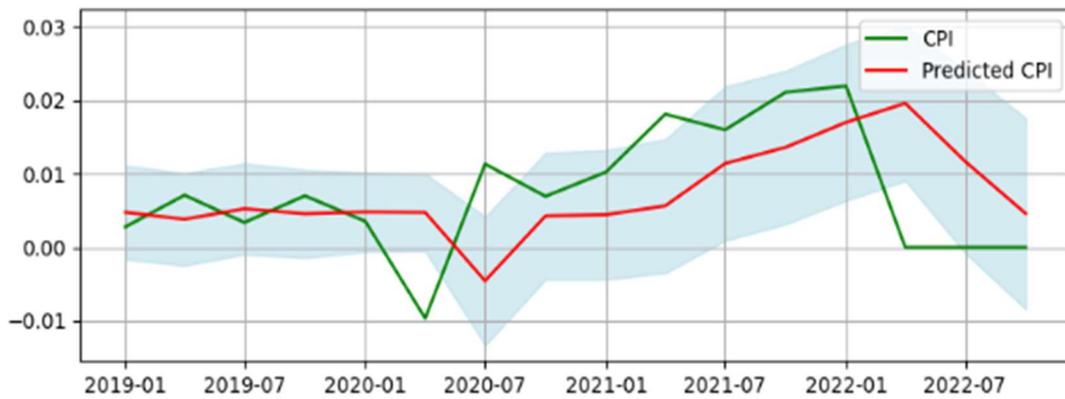


Figura 10 - Previsão segundo o ARIMA(0,1,1) para o CPI

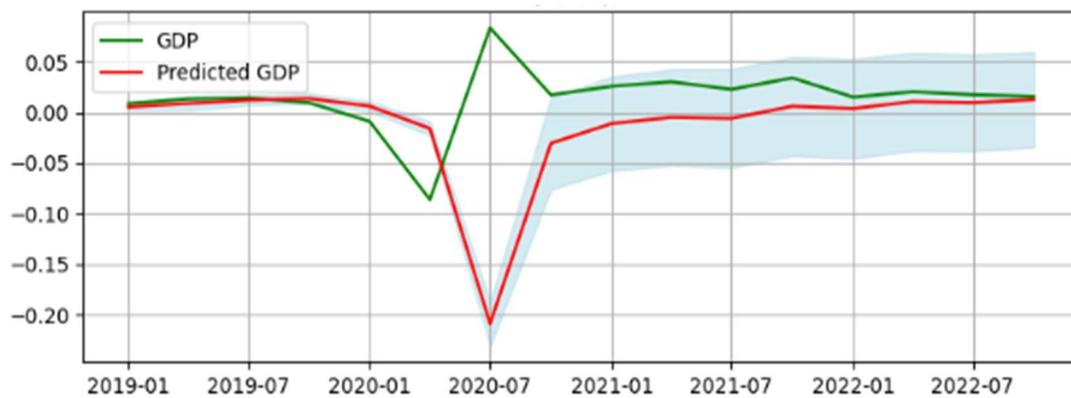


Figura 11 - Previsão segundo o ARIMA(1,1,3) para o GDP

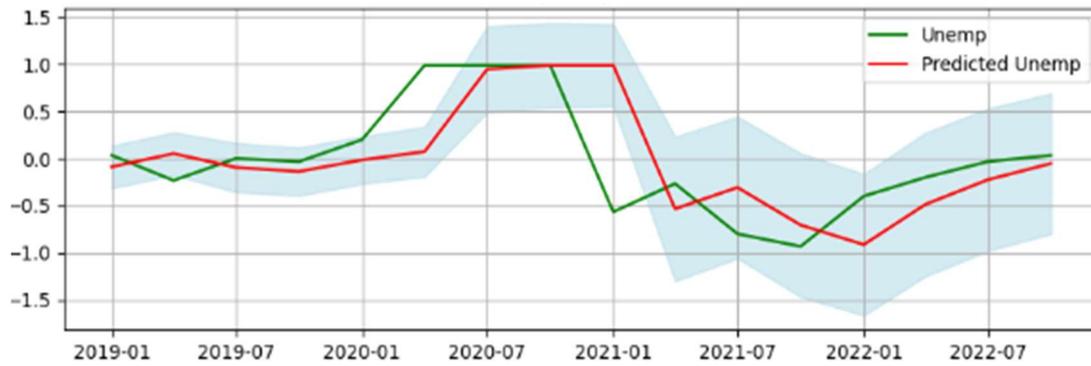


Figura 12 - Previsão segundo o ARIMA(1,1,0) para o Unemployment

3.2.1.1. Análise dos Resíduos para os Modelos Univariados

Todos os pressupostos dos resíduos têm de ser satisfeitos, caso contrário o modelo não será válido para a previsão das séries em estudo.

O primeiro pressuposto a estudar é se os resíduos são independentes através do teste de Durbin-Watson.

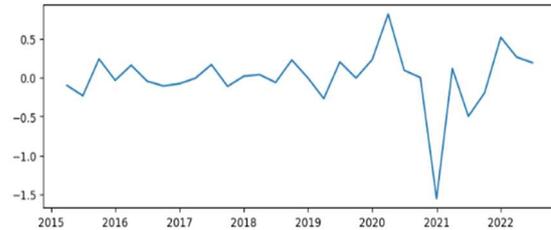
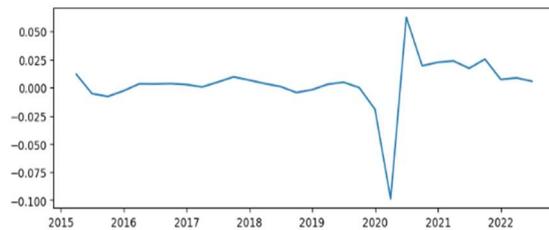
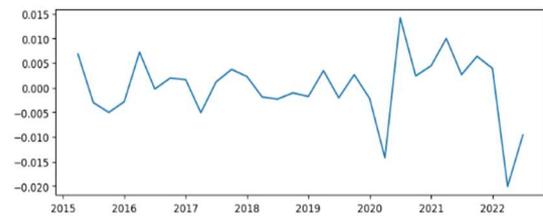
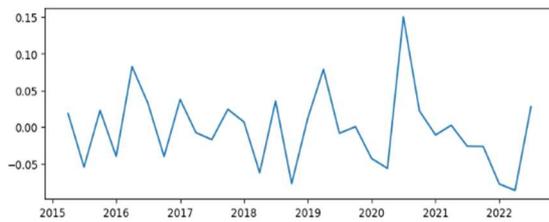
Variável	Teste de Durbin-Watson
S&P 500	2.22
CPI	1.79
GDP	2.08
Unemployment Rate	1.95

Segundo o teste de Durbin-Watson, todas as variáveis são independentes uma vez que o valor está perto de 2.

O segundo pressuposto a estudar é se os resíduos têm valor médio nulo através das estatísticas descritivas das mesmas:

Variável	Valor Médio
S&P 500	0.00
CPI	0.00
GDP	0.00
Unemployment Rate	0.00

O terceiro e último pressuposto é verificar se existem efeitos ARCH nos resíduos uma vez que estes indicam a existência de heterocedasticidade: variância não constante. Vamos realizar a representação gráfica dos resíduos das quatro séries.



Como constatado, confirmados todos os pressupostos dos resíduos, o próximo passo será verificar a performance de cada um dos modelos univariados através do auxílio de medidas de erro, como o MSE.

3.2.1.2. Previsão das variáveis

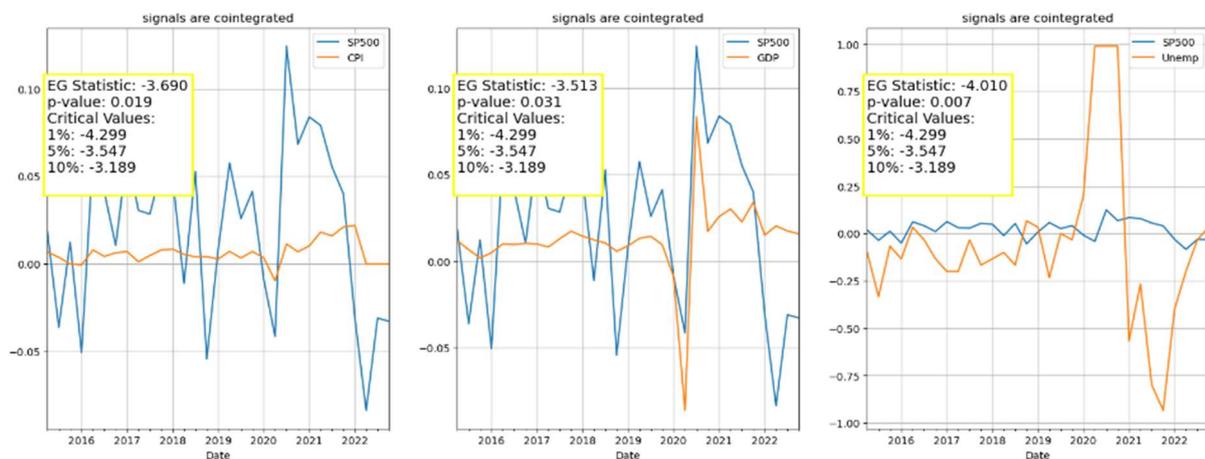
Uma vez que, graficamente, as previsões realizadas pelos modelos univariados para as quatro variáveis não apresentam a melhor previsão, devemos ainda assim olhar para o MSE.

Variável	MSE
S&P 500	0.00
CPI	0.00
GDP	0.00
Unemployment Rate	0.00

3.2.2. Estimação dos Modelos Multivariados

É também importante utilizar modelos multivariados na previsão das variáveis em estudo sempre que pretendemos verificar a relação entre uma ou mais variáveis embora não haja dependência estatística. Pode haver contexto em que as duas variáveis estão relacionadas de forma mais complexa e análises multivariadas podem ser justificadas. No entanto, nenhuma das variáveis parece apresentar linearidade nas suas relações com as restantes tornando a estimação do modelo multivariado complexa. Contudo, foram realizados vários testes por forma a alcançar uma previsão credível e relevante para o estudo, mas sem sucesso.

Da tentativa de estimação do modelo multivariado, retira-se, para o presente estudo, a análise da cointegração das variáveis, de acordo com o Teste do Engle-Granger.



A cointegração é uma relação estatística entre duas ou mais séries temporais que têm uma relação de longo prazo, mesmo que sejam não estacionárias individualmente. Isso significa que, embora cada série temporal possa ser não estacionária e tenha tendências individuais, uma combinação linear delas pode ser estacionária.

Foi realizado o teste de Engle-Granger para os pares de variáveis (S&P 500, CPI), (S&P 500, GDP) e (S&P 500, Unemployment Rate) uma vez que a variável “dependente” da movimentação das restantes é a S&P 500, é possível verificar o acompanhamento não-linear por parte desta na evolução das restantes: as variáveis independentes aumentam, o S&P 500 parece aumentar; as variáveis independentes diminuem, o S&P 500 tem tendência para descer.

No entanto, estas relações não sendo lineares e obtendo *spikes* na sua curva de movimentação, torna complexa a parametrização de um modelo multivariado e que cumpra todos os pressupostos exigidos.

CAPÍTULO 4

Conclusões

A relação entre a inflação e o S&P 500 é influenciada por uma série de fatores e pode variar ao longo do tempo. A inflação pode afetar as ações de empresas que compõem o índice de diferentes maneiras, dependendo das circunstâncias económicas, das políticas do Fed, e das expectativas do mercado.

O objetivo desta dissertação passou por tentar relacionar as variáveis que refletem a inflação com o crescimento do S&P 500 e tentar prever, utilizando métodos adequados de modelação.

Para executar as previsões das nossas séries temporais foi utilizado um tipo de modelo econométrico univariado: o modelo ARIMA, dado que as variáveis não apresentam comportamentos lineares que através de um modelo multivariado levariam a um desvio notável do registado até ao momento.

Para as séries S&P 500, CPI e GDP foi necessário aplicar as diferenças logaritmizadas para chegar à estacionaridade e ao Unemployment Rate aplicou-se a diferenciação usual. Assim, chegou-se à estacionaridade após uma única diferenciação.

CAPÍTULO 5

Referências Bibliográficas

Blinder, A S (1982), "Chapter 12: The Anatomy of Double-Digit Inflation in the 1970s", in R E Hall (ed.), *Inflation: Causes and Effects*, University of Chicago Press, p.261-282.

Coins News (2023) https://www.usinflationcalculator.com/inflation/inflation-vs-consumer-price-index-cpi-how-they-are-different/#google_vignette

D.T. Anderson, J.C. Bezdek, M. Popescu, and J.M. Keller (2010), "Comparing Fuzzy, Probabilistic, and Possibilistic Partitions", *IEEE Transactions on Fuzzy Systems*, 18(5), 906-918.

J.B. Maverick (2023) "S&P 500 Average Return" Article in Investopedia
<https://www.investopedia.com/s-and-p-500-little-changed-after-inflation-data-7969457>
<https://www.spglobal.com/marketintelligence/en/news-insights/blog/insurers-get-to-grips-with-evolving-net-zero-standards>

Jongrim Ha, M. Ayhan Kose, and Franziska Ohnsorge (2021) "Inflation During the Pandemic: What Happened? What is Next?" in World Bank

Ha, Jongrim; Ivanova, Anna; Ohnsorge, Franziska; Unsal, Filiz. (2019) "Inflation: Concepts, Evolution, and Correlates" Policy Research Working Paper;No. 8738. © World Bank, Washington, DC.

Hiranmayi Srinivasan (2023) "U.S. Inflation Rate by Year: 1929–2023" Article in Investopedia

H. Akaike (1973), "Information Theory as an Extension of the Maximum Likelihood Principle", in B. N. Petrov, and F. Csaki, (Eds.), *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, pp. 267-281.

Gagnon, J E (2022), "Why US Inflation Surged in 2021 and What the Fed Should Do to Control It", PIIIE series on inflation, 11 March 11, Peterson Institute for International Economics.

Maram Ibrahim (2019), "Inflation" at Research Gate website
(<https://www.researchgate.net/publication/332763097>)

Olivier Armantier, Leo Goldman, Gizem Kosar, Giorgio Topa, Wilbert van der Klaauw, and John C. Williams (2022), "What Are Consumers' Inflation Expectations Telling Us Today?," Federal Reserve Bank of New York Liberty Street Economics

U.S. Bureau of Labor Statistics. "Consumer Price Index."

U.S. Bureau of Economic Analysis. "Gross Domestic Product."

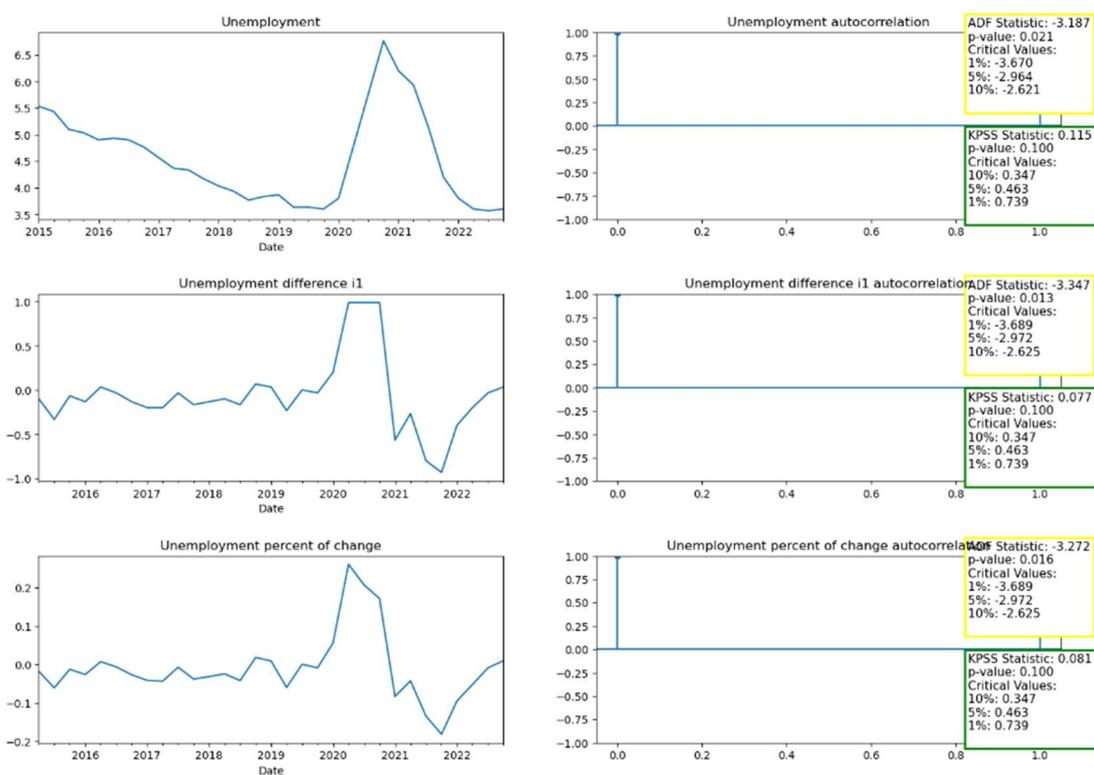
U.S. Bureau of Labor Statistics. "Consumer Price Index—December 2021."

U.S. Bureau of Labor Statistics. "Consumer Price Index—December 2022."

Anexos

	SP500	CPI	GDP	Unemp
count	32.000000	32.000000	32.000000	32.000000
mean	2980.581314	255.449375	21193.879219	4.546875
std	816.694649	15.520269	2417.446479	0.865990
min	1951.224918	235.355000	18063.529000	3.566667
25%	2290.665570	243.412667	19232.407750	3.800000
50%	2791.539075	253.062500	20858.298500	4.350000
75%	3628.869545	261.717000	22168.422750	5.050000
max	4602.108906	284.893667	26408.405000	6.766667

Estatística descritiva das séries em estudo



Teste ADF e KPSS para a série Unemployment

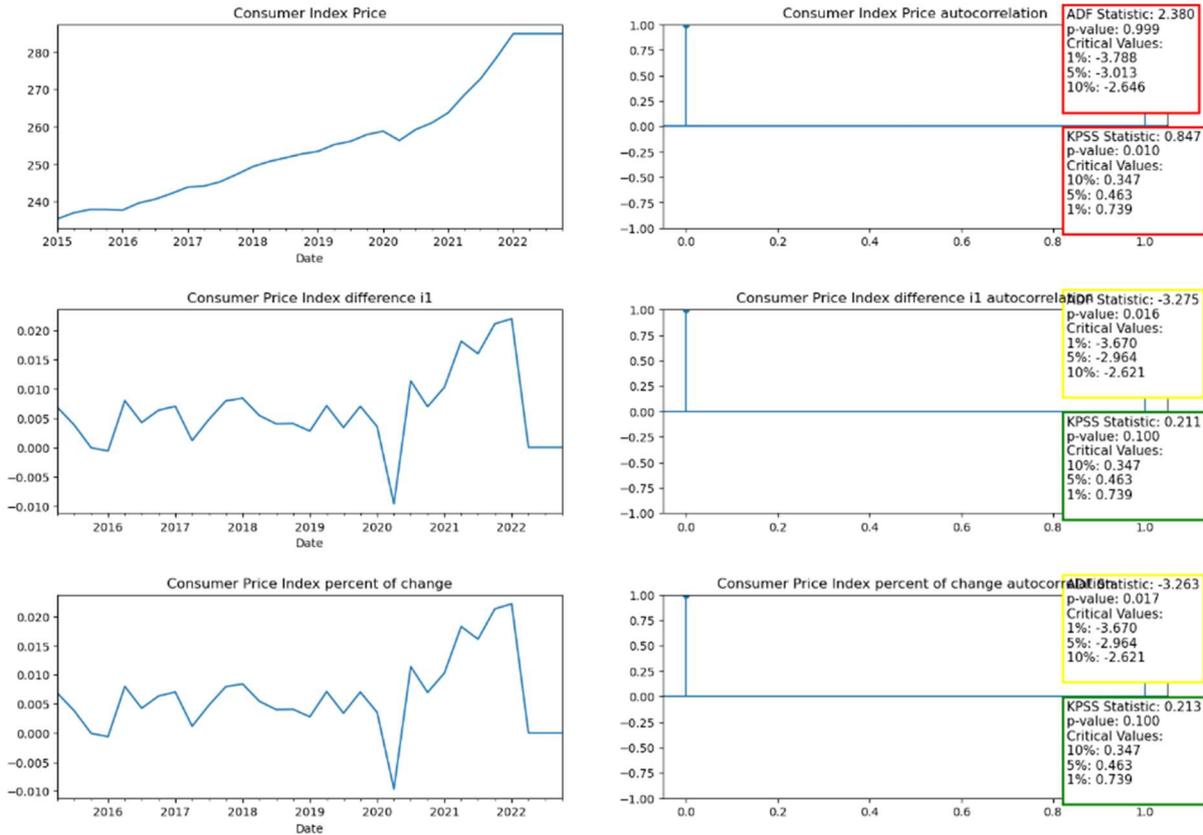
Phillips-Perron Test (Z-tau)

```

=====
Test Statistic      -1.440
P-value             0.849
Lags                10
-----
    
```

Trend: Constant and Linear Time Trend
 Critical Values: -4.28 (1%), -3.56 (5%), -3.22 (10%)
 Null Hypothesis: The process contains a unit root.
 Alternative Hypothesis: The process is weakly stationary.

Teste PP para a série Unemployment



Teste ADF e KPSS para a série CPI

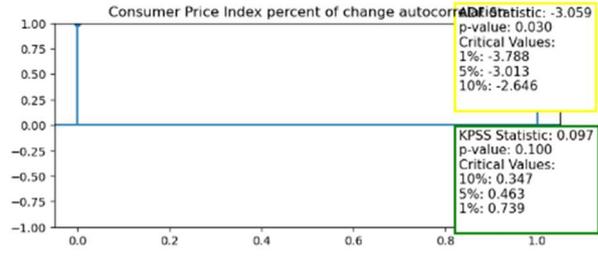
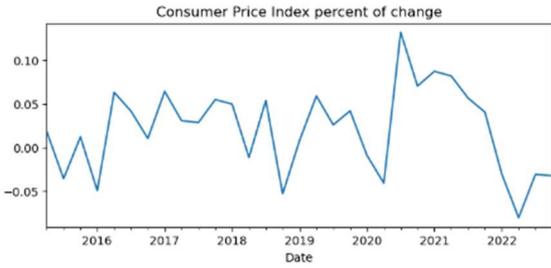
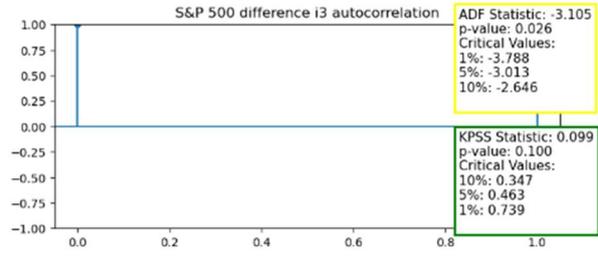
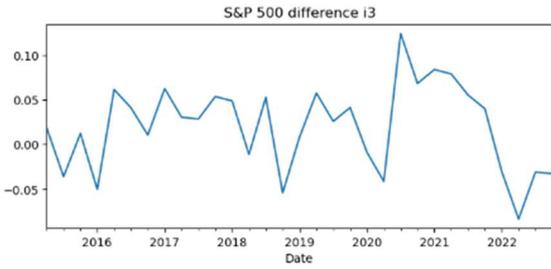
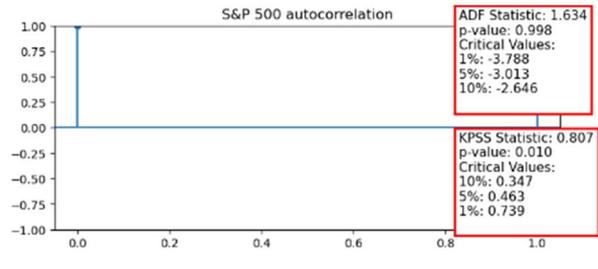
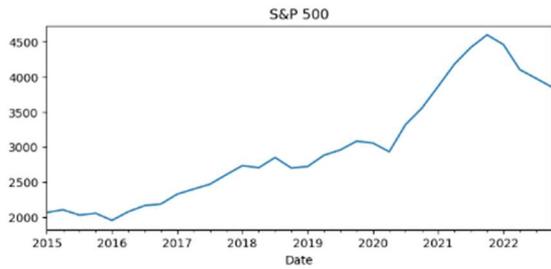
Phillips-Perron Test (Z-tau)

```

=====
Test Statistic      -0.996
P-value             0.945
Lags                10
-----
    
```

Trend: Constant and Linear Time Trend
 Critical Values: -4.28 (1%), -3.56 (5%), -3.22 (10%)
 Null Hypothesis: The process contains a unit root.
 Alternative Hypothesis: The process is weakly stationary.

Teste PP para a série CPI



Teste ADF e KPSS para a série S&P 500

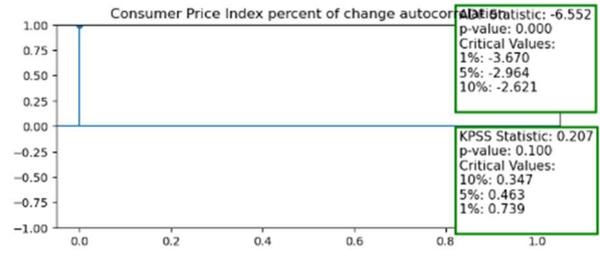
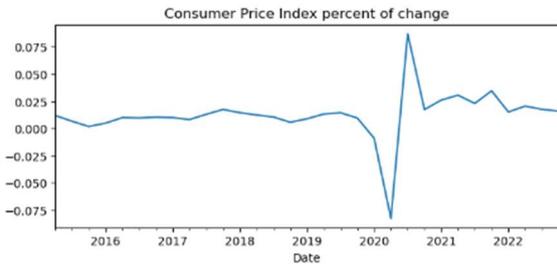
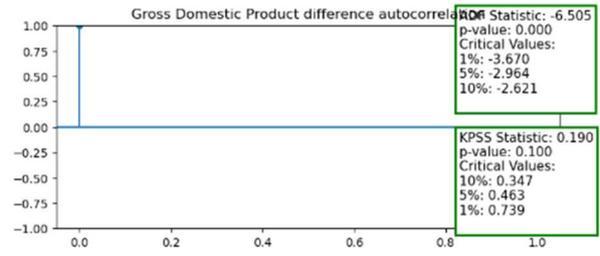
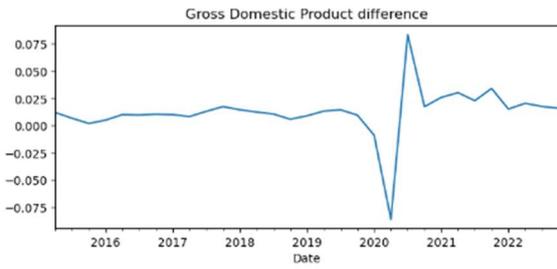
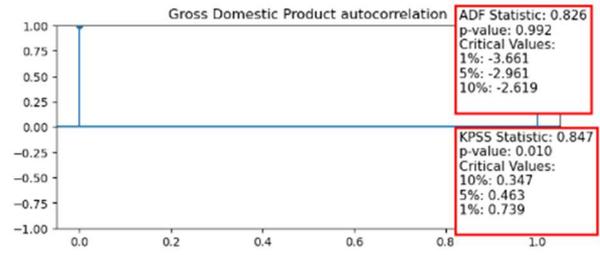
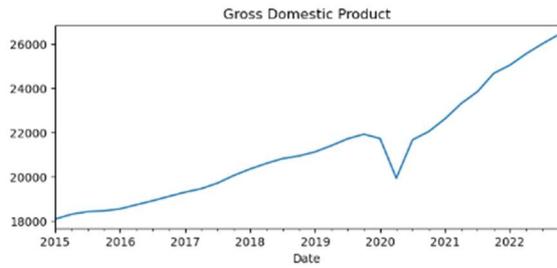
```

Phillips-Perron Test (Z-tau)
=====
Test Statistic          -1.411
P-value                  0.858
Lags                     10
-----

```

Trend: Constant and Linear Time Trend
Critical Values: -4.28 (1%), -3.56 (5%), -3.22 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.

Teste PP para a série S&P 500



Teste ADF e KPSS para a série GDP

```

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=-38.801, Time=0.03 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=-47.848, Time=0.07 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=-46.597, Time=0.10 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=-40.799, Time=0.03 sec
ARIMA(2,1,0)(0,0,0)[0] intercept : AIC=-47.328, Time=0.13 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=-46.804, Time=0.07 sec
ARIMA(2,1,1)(0,0,0)[0] intercept : AIC=-45.445, Time=0.14 sec
ARIMA(1,1,0)(0,0,0)[0] : AIC=-49.823, Time=0.04 sec
ARIMA(2,1,0)(0,0,0)[0] : AIC=-49.318, Time=0.06 sec
ARIMA(1,1,1)(0,0,0)[0] : AIC=-48.795, Time=0.06 sec
ARIMA(0,1,1)(0,0,0)[0] : AIC=-48.517, Time=0.06 sec
ARIMA(2,1,1)(0,0,0)[0] : AIC=-47.428, Time=0.09 sec

```

```

Best model: ARIMA(1,1,0)(0,0,0)[0]
Total fit time: 0.888 seconds

```

SARIMAX Results

Dep. Variable:	y	No. Observations:	16
Model:	SARIMAX(1, 1, 0)	Log Likelihood	26.911
Date:	Tue, 31 Oct 2023	AIC	-49.823
Time:	16:25:11	BIC	-48.407
Sample:	04-01-2015	HQIC	-49.838
	- 01-01-2019		
Covariance Type:	opg		
	coef	std err	z P> z [0.025 0.975]
ar.L1	-0.7390	0.162	-4.556 0.000 -1.057 -0.421
sigma2	0.0015	0.001	1.951 0.051 -7.17e-06 0.003
Ljung-Box (L1) (Q):	0.85	Jarque-Bera (JB):	0.62
Prob(Q):	0.36	Prob(JB):	0.73
Heteroskedasticity (H):	0.82	Skew:	0.14
Prob(H) (two-sided):	0.83	Kurtosis:	2.05

Aplicação do *pm.autoarima* na série S&P 500 no dataset de treino

Performing stepwise search to minimize aic

```

ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=-113.580, Time=0.08 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=-112.202, Time=0.09 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=-116.495, Time=0.14 sec
ARIMA(0,1,0)(0,0,0)[0]          : AIC=-115.540, Time=0.05 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=-113.371, Time=0.20 sec
ARIMA(0,1,2)(0,0,0)[0] intercept : AIC=-114.700, Time=0.24 sec
ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=-113.841, Time=0.33 sec
ARIMA(0,1,1)(0,0,0)[0]          : AIC=-116.247, Time=0.06 sec

```

Best model: ARIMA(0,1,1)(0,0,0)[0] intercept
 Total fit time: 1.180 seconds

SARIMAX Results

Dep. Variable:	y	No. Observations:	15			
Model:	SARIMAX(0, 1, 1)	Log Likelihood	61.247			
Date:	Tue, 31 Oct 2023	AIC	-116.495			
Time:	15:52:54	BIC	-114.578			
Sample:	04-01-2015	HQIC	-116.672			
	- 10-01-2018					
Covariance Type:	opg					
	coef	std err	z	P> z 	[0.025	0.975]
intercept	0.0001	0.000	0.591	0.555	-0.000	0.001
ma.L1	-0.9887	6.036	-0.164	0.870	-12.818	10.841
sigma2	7.706e-06	4.6e-05	0.167	0.867	-8.25e-05	9.79e-05
Ljung-Box (L1) (Q):	0.54	Jarque-Bera (JB):	0.80			
Prob(Q):	0.46	Prob(JB):	0.67			
Heteroskedasticity (H):	0.37	Skew:	0.06			
Prob(H) (two-sided):	0.30	Kurtosis:	1.83			

Aplicação do *pm.autoarima* na série CPI do set de treino

```

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=-114.850, Time=0.10 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=-114.126, Time=0.05 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=inf, Time=0.14 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=-116.624, Time=0.03 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=-116.815, Time=0.13 sec
ARIMA(2,1,1)(0,0,0)[0] intercept : AIC=-115.227, Time=0.20 sec
ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=-116.951, Time=0.28 sec
ARIMA(0,1,2)(0,0,0)[0] intercept : AIC=inf, Time=0.20 sec
ARIMA(2,1,2)(0,0,0)[0] intercept : AIC=-115.363, Time=0.27 sec
ARIMA(1,1,3)(0,0,0)[0] intercept : AIC=-115.290, Time=0.34 sec
ARIMA(0,1,3)(0,0,0)[0] intercept : AIC=inf, Time=0.16 sec
ARIMA(2,1,3)(0,0,0)[0] intercept : AIC=-107.997, Time=0.30 sec
ARIMA(1,1,2)(0,0,0)[0] : AIC=-118.205, Time=0.14 sec
ARIMA(0,1,2)(0,0,0)[0] : AIC=-118.765, Time=0.12 sec
ARIMA(0,1,1)(0,0,0)[0] : AIC=-118.637, Time=0.07 sec
ARIMA(0,1,3)(0,0,0)[0] : AIC=inf, Time=0.12 sec
ARIMA(1,1,1)(0,0,0)[0] : AIC=-117.209, Time=0.06 sec
ARIMA(1,1,3)(0,0,0)[0] : AIC=-120.568, Time=0.23 sec
ARIMA(2,1,3)(0,0,0)[0] : AIC=-106.625, Time=0.23 sec
ARIMA(1,1,4)(0,0,0)[0] : AIC=-106.624, Time=0.24 sec
ARIMA(0,1,4)(0,0,0)[0] : AIC=-114.306, Time=0.27 sec
ARIMA(2,1,2)(0,0,0)[0] : AIC=-117.235, Time=0.31 sec
ARIMA(2,1,4)(0,0,0)[0] : AIC=-113.576, Time=0.29 sec

```

Best model: ARIMA(1,1,3)(0,0,0)[0]
Total fit time: 4.309 seconds

SARIMAX Results

Dep. Variable:	y	No. Observations:	15			
Model:	SARIMAX(1, 1, 3)	Log Likelihood	65.284			
Date:	Tue, 31 Oct 2023	AIC	-120.568			
Time:	15:53:00	BIC	-117.372			
Sample:	04-01-2015	HQIC	-120.863			
	- 10-01-2018					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.0495	0.737	0.067	0.946	-1.394	1.494
ma.L1	0.8051	2.187	0.368	0.713	-3.481	5.092
ma.L2	-0.9369	3.837	-0.244	0.807	-8.457	6.583
ma.L3	-0.8053	1.895	-0.425	0.671	-4.519	2.908
sigma2	3.485e-06	8.18e-06	0.426	0.670	-1.26e-05	1.95e-05
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	1.13			
Prob(Q):	0.99	Prob(JB):	0.57			
Heteroskedasticity (H):	0.85	Skew:	-0.09			
Prob(H) (two-sided):	0.86	Kurtosis:	1.62			

Aplicação do *pm.autoarima* na série GDP do set de treino

```

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=-10.971, Time=0.04 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=-14.742, Time=0.06 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=inf, Time=0.18 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=-12.872, Time=0.02 sec
ARIMA(2,1,0)(0,0,0)[0] intercept : AIC=-12.742, Time=0.07 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=-12.742, Time=0.09 sec
ARIMA(2,1,1)(0,0,0)[0] intercept : AIC=-13.101, Time=0.14 sec
ARIMA(1,1,0)(0,0,0)[0] : AIC=-16.285, Time=0.07 sec
ARIMA(2,1,0)(0,0,0)[0] : AIC=-14.286, Time=0.05 sec
ARIMA(1,1,1)(0,0,0)[0] : AIC=-14.286, Time=0.05 sec
ARIMA(0,1,1)(0,0,0)[0] : AIC=inf, Time=0.06 sec
ARIMA(2,1,1)(0,0,0)[0] : AIC=inf, Time=0.13 sec

```

Best model: ARIMA(1,1,0)(0,0,0)[0]
Total fit time: 0.962 seconds

SARIMAX Results

Dep. Variable:	y	No. Observations:	15
Model:	SARIMAX(1, 1, 0)	Log Likelihood	10.143
Date:	Tue, 31 Oct 2023	AIC	-16.285
Time:	15:53:05	BIC	-15.007
Sample:	04-01-2015	HQIC	-16.403
	- 10-01-2018		

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.6671	0.271	-2.462	0.014	-1.198	-0.136
sigma2	0.0132	0.009	1.549	0.121	-0.003	0.030

Ljung-Box (L1) (Q): 0.01 Jarque-Bera (JB): 0.91

Prob(Q): 0.91 Prob(JB): 0.63

Heteroskedasticity (H): 0.62 Skew: -0.09

Prob(H) (two-sided): 0.61 Kurtosis: 1.76

Aplicação do *pm.autoarima* na série de Unemployment do set de treino