

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

U LISBOA

UNIVERSIDADE
DE LISBOA

Análise de sensibilidade: Fatores que influenciam o preço das Criptomoedas

Emine Ahmedova Karameshinova

Mestrado em Matemática Financeira

Orientadora:

Professora Doutora, Diana Elisabeta Aldea Mendes, Professora Associada, Iscte-Iul

Outubro, 2023

Departamento de Finanças

Departamento de Matemática

Análise de sensibilidade: Fatores que influenciam o preço das Criptomoedas

Emine Ahmedova Karameshinova

Mestrado em Matemática Financeira

Orientadora:

Professora Doutora, Diana Elisabeta Aldea Mendes, Professora
Associada, Iscte-Iul

Outubro, 2023

“Once you know what failure feels like, determination chases success.”

- Kobe Bryant

Agradecimento

Sempre me disseram que a dissertação de mestrado era uma longa e difícil caminhada, durante a qual experienciamos uma mistura de sentimentos e emoções, de momentos altos e baixo. A minha caminhada foi permeada por inúmeros desafios, tristezas, incertezas e percalços, mas no fim são os sentimentos de orgulho e realização que ecoam por dentro.

Fazendo uma retrospectiva do percurso que hoje vê o seu fim, percebo que nunca estive só. Alcanço mais um objetivo com o sentimento de gratidão para aqueles que me acompanharam e a cujas palavras que se seguem dedico.

À minha orientadora Diana Mendes, endereço o meu mais sincero agradecimento por me ter instruído e orientado na elaboração desta dissertação. A partilha do seu conhecimento, as críticas construtivas e sugestões foram fundamentais para o resultado desta investigação.

Ainda a nível académico, não posso deixar de agradecer à instituição que me acolheu e deu tantas oportunidades, o Iscte – Instituto Universitário de Lisboa, que se tornou Casa nos últimos anos. Aqui sei que cresci e aprendi muito a nível académico, pessoal e profissional. Na certeza de que este não será um adeus, mas sim um até já.

Aos meus amigos e colegas de trabalho, sem enumerar na possibilidade de me faltar alguém injustificadamente, por me incentivarem a continuar e por estarem a meu lado nesta longa caminhada que é a vida.

À pessoa mais especial para a qual as palavras não chegam, Henrique Bota, por ser o meu porto de abrigo, a minha motivação e o meu apoio incondicional. À pessoa que nunca me deixou desistir, por confiar mais em mim do que eu própria, e para a qual profiro a seguinte palavra com a certeza de que nunca foi tão verdadeira: Obrigada.

Por fim, o meu agradecimento mais profundo é, sem dúvida, para os meus pais, que são as pessoas graças às quais escrevo estas palavras. Apesar de tudo, nunca me faltou apoio, incentivo e motivação da parte deles. Sempre lutaram contra todas as adversidades para me proporcionar o melhor percurso possível pelo qual serei eternamente grata.

*“Agora que eu tenho a responsabilidade
E há coisas que eu não vou poder arriscar
Mas tenho de continuar a sonhar”*

Resumo

O mercado de criptomoedas tem vindo a crescer exponencialmente desde 2009, com o surgimento da *Bitcoin*. Dadas as suas características de segurança, descentralização e imutabilidade, bem como a sua forte volatilidade, esta tem sido alvo de grande mediatismo e estudo.

A presente dissertação tem como objetivo perceber o funcionamento das criptomoedas para, posteriormente, aplicar modelação preditiva e entender como é que os fatores Preço da *Ethereum*, Preço do Ouro, Preço do Petróleo, Índice acionista S&P500 e VIX influenciam o preço da *Bitcoin*.

Para o efeito, recorreremos à família de modelos *Auto Regressive Integrated Moving Average* (ARIMA) e aplicámos à série temporal da *Bitcoin* entre 03 de junho de 2019 e 31 de maio de 2023. Também foi estudada a correlação entre as variáveis e testada a causalidade de Granger.

Em face dos resultados obtidos concluímos, primeiramente, que a biblioteca automatizada de estimação do modelo – função *auto_arima()* – não consegue capturar corretamente os efeitos das variáveis exógenas, razão pela qual a estimação do modelo final a aplicar é concretizada de forma manual.

Desde modo, inferimos que o modelo que melhor se ajusta aos dados e com melhor capacidade preditiva é o ARIMA(4,1,2) e, mais concretamente, após estudo de diferentes combinações das variáveis exógenas, o ARIMAX(4,1,2) com a inclusão da variável *Ethereum*, seguindo-se da variável Ouro.

Em suma, concluímos que o preço da *Bitcoin* é influenciado positivamente pela variável *Ethereum* e pela variável Ouro, embora mais fracamente. Por outro lado, as variáveis SP500, Petróleo e VIX não mostram ter qualquer impacto significativo.

Palavras-chave: Criptomoedas, *Bitcoin*, Preço, ARIMA, Previsão.

Abstract

The cryptocurrency market has been growing exponentially since 2009 with the emergence of Bitcoin. Given to its characteristics of security, decentralization and immutability, as well as its strong volatility, it has been the subject of great media and studies.

This dissertation aims to understand how cryptocurrencies work in order to apply predictive modeling and understand how the factors Ethereum Price, Gold Price, Oil Price, S&P500 Stock Index and VIX influence the price of Bitcoin.

To do this, we applied the family models of *Auto Regressive Integrated Moving Average* (ARIMA) to the Bitcoin time series between June 3rd, 2019 and May 31st, 2023. The correlation between the variables was also studied and Granger's causality was tested.

In view of the results obtained, we concluded that the automated model estimation library – function *auto_arima* () – cannot correctly capture the effects of exogenous variables, which is why the estimation of the final model to be applied is performed manually.

Thus, we infer that the model that best fits the data, while have greater predictive capacity is the ARIMA(4,1,2) and, more specifically, after studying different combinations of exogenous variables, the ARIMAX(4,1,2) with the inclusion of the Ethereum variable, followed by the Gold variable.

In short, we concluded that the price of Bitcoin is positively influenced by the Ethereum variable and the Gold variable, although more weakly. On the other hand, the SP500, Oil, and VIX variables do not show any significant impact.

Keywords: Cryptocurrencies, Bitcoin, Price, ARIMA, Prediction.

Índice

Agradecimento	iii
Resumo	v
<i>Abstract</i>	vii
Capítulo 1. Introdução	1
Capítulo 2. Revisão da Literatura	3
2.1. As Criptomoedas	3
2.2. Preço das Criptomoedas e fatores que influenciam	6
Capítulo 3. Metodologia	9
3.1. Estacionariedade	9
3.2. Causalidade de Granger	10
3.3. Modelo ARIMA	11
3.4. Previsão e Erro de Previsão	13
Capítulo 4. Dados e análise de estatísticas descritivas dos dados	15
Capítulo 5. Resultados e Discussão	19
5.1. Estacionariedade das séries	19
5.2. Estimação do modelo e previsão	20
Capítulo 6. Conclusão e recomendações	25
Referências Bibliográficas	27
Anexo A	29
Anexo B	30
Anexo C	33

Índice de figuras

Figura 1. Representação gráfica das variáveis Ouro, Petróleo, SP500 e VIX.	15
Figura 2. Representação gráfica do preço da <i>Bitcoin</i> e <i>Ethereum</i> .	16
Figura 3. Mapa de calor de correlação das variáveis.	19
Figura 4. Previsão de 10 semanas com o modelo ARIMA, para vários lags na parte AR.	20
Figura 5. Representação gráfica do conjunto de teste e valores preditos para o Modelos ARIMA e ARIMAX(4,1,2).	23

Índice de tabelas

Tabela 1. Estatísticas descritivas das séries temporais.	17
Tabela 2. Teste de causalidade de Granger.	18
Tabela 3. P-values dos testes de estacionariedade para as séries temporais em níveis.	19
Tabela 4. P-values dos testes de estacionariedade para as primeiras diferenças logaritmizadas das séries temporais.	20
Tabela 5. Valores do critério de informação AIC e métricas de erro MAPE e RMSE dos modelos ARIMA(2,1,2) e ARIMAX(0,1,1).	23

CAPÍTULO 1

Introdução

Ao longo da história da Humanidade, o conceito de moeda tem evoluído de diferentes formas e com a revolução digital e da *internet*, a evolução da moeda física para a moeda digital foi inevitável. Pela primeira vez na história, graças aos recentes avanços na tecnologia *Blockchain*, foi possível transferir valor de qualquer ponto do mundo sem o uso de um intermediário, como por exemplo, um banco ou uma bolsa. A este fenómeno deu-se o nome de Criptomoeda.

Uma característica notável das criptomoedas é a volatilidade inerente aos preços. Em curtos espaços de tempo, os preços destes ativos digitais podem sofrer variações diferenciadas, resultando em lucros substanciais ou em riscos de perda significativas. Com a crescente importância das criptomoedas nos mercados financeiros, a compreensão dos fatores que influenciam este comportamento dos preços torna-se uma questão de extrema relevância.

Atualmente, o mercado de criptomoedas tem uma extensão inimaginável no tempo quando apareceu a primeira, a *Bitcoin*, em 2009. Passado quase cerca de 15 anos da sua criação, a *Bitcoin* é a criptomoeda líder do mercado, com um preço de mais de 34 mil dólares a unidade¹.

Neste sentido, a questão principal desta dissertação trata em perceber como é que o preço da *Bitcoin* se comporta e se existem variáveis financeiras exógenas que causam a *Bitcoin*, aplicando modelação preditiva através de modelos *Auto Regressive Integrated Moving Average* (ARIMA), verificar a acurácia com que conseguimos prever os seus valores.

A presente dissertação encontra-se estruturada em seis capítulos sendo o primeiro referente à Introdução, expressa anteriormente.

No segundo capítulo é feita uma breve revisão da literatura sobre o tema em estudo, composto por duas partes, sendo na primeira apresentada um estudo teórico à história das criptomoedas, referindo alguns conceitos associados e a origem, bem como uma breve explicação do seu funcionamento, com referência em específico à *Bitcoin* e *Ethereum*. Na segunda parte é feito um estudo aprofundado da literatura no que concerne aos fatores que influenciam o preço das criptomoedas e, em particular, o preço da *Bitcoin*.

No terceiro capítulo, concentramo-nos nos conceitos gerais de análise de séries temporais e nas diferentes metodologias e testes para concretizar o objetivo. Deste modo, dividimos o capítulo em quatro partes: abordagem de conceitos gerais de econometria, tais como raiz unitária e

estacionariedade; referência ao teste de causalidade de Granger; explicação dos modelos utilizado para alcançar os resultados desta investigação e critérios para a escolha do modelo final.

No quarto capítulo são identificadas e descritas as séries temporais utilizadas no estudo empírico, acompanhadas de uma breve análise de algumas estatísticas descritivas das mesmas e do estudo da estacionariedade.

No quinto capítulo será aplicada a metodologia identificada e feita a comparação dos modelos obtidos, bem como a apresentação e discussão dos principais resultados para auferir as relações existentes entre o preço da *Bitcoin* e as variáveis exógenas consideradas.

Por último, no sexto capítulo, é apresentada a conclusão, onde são tiradas as ilações finais, bem como pistas para futuras investigações, sempre profícuas para o desenvolvimento científico da temática.

Revisão da Literatura

2.1. As Criptomoedas

A era digital incitada nas últimas décadas introduziu muitas tecnologias de informação e comunicação que vieram a mudar completamente a velocidade e forma de partilha de informação. Hoje, a informação está a moldar rapidamente as nossas expectativas e formas de ver o mundo.

Do ponto de vista financeiro, o acesso rápido e fácil à informação e o menor custo associado aumentaram o seu impacto e envolvimento nas nossas decisões de investimento (Gunay, 2019). Com isto, mudaram também os hábitos de consumo e introduziram-se novos produtos.

O interesse pelo mercado de criptomoedas tem aumentado exponencialmente e, apesar da invenção recente, a utilização das mesmas tem proporcionado uma reviravolta no sistema monetário tradicional (Farell, 2015). Como exemplo disso temos os *hedge funds* e as empresas de gestão de ativos que passaram a incluir este tipo de ativos nas suas carteiras.

Quando falamos de criptomoedas, falamos de uma família de tipos de moedas digitais diferentes e mais, quando falamos de cripto ativos falamos de uma série de ativos distintos, dos quais as criptomoedas são apenas uma componente (Mónica, 2022).

Não existe uma definição exata e aceite por todos relacionada com as criptomoedas. O conceito de criptomoeda foi inicialmente introduzido por Wei Dai, em 1998, com a proposta da sua versão de moeda virtual chamada *B-money* – um sistema de dinheiro digital anónimo e distribuído – que segundo o mesmo “permite que entidades com pseudónimos não rastreáveis cooperem entre si de forma mais eficiente, fornecendo-lhes um meio de troca e um método de cumprimento de contratos”.

Em 2012, o Banco Central Europeu (BCE) definiu moeda virtual como "um tipo de dinheiro digital não regulamentado, que é emitido e normalmente controlado pelos seus criadores, e utilizado e aceite entre os membros de uma comunidade virtual específica". Na categoria de moeda virtual estão inseridas as criptomoedas que utilizam criptografia como principal fonte de segurança.

De um modo simples, a criptomoeda – ou moeda digital – é um meio de troca que funciona como o dinheiro (na medida em que pode ser trocado por bens e serviços), mas, ao contrário da moeda tradicional, não está vinculada e é independente de qualquer autoridade central ou entidade

governamental (Maese *et al.*, 2016). Por outras palavras, é um sistema de troca descentralizado *peer-to-peer* (termo em inglês que significa “pessoa para pessoa”) que permite aos utilizadores efetuar pagamentos de bens e serviços sem terem a necessidade de regularização por parte de uma autoridade central. Assim, estas dependem da transmissão de informação digital, utilizando métodos criptográficos para garantir transações legítimas e únicas (Doran, 2014).

A tecnologia que está na base das criptomoedas e que permite as mesmas terem as características já mencionadas é a *Blockchain*, que funciona como um livro-razão distribuído e público e que permite armazenar cronologicamente toda a informação inerente às mesmas (Mattila, 2016; Pacheco, 2018).

O que torna esta tecnologia tão importante é, principalmente, o seu protocolo de segurança que é garantido graças à descentralização, funcionando através de um esquema *peer-to-peer* de computadores distribuídos globalmente (*nodes*) e, por isso, sem a necessidade de intervenção de um intermediário (Rodrigues, 2019). Neste sentido, qualquer pessoa pode fazer parte integrante da *Blockchain* e, por ser público, tem acesso a todo o historial das transações registadas até ao momento (Pacheco, 2018).

De uma forma resumida e segundo Mónica (2022), a *Blockchain* tem uma série de características de segurança e vantagens muito importantes, como sejam, imutabilidade, transparência, segurança, estabilidade e eficiência.

A primeira criptomoeda surgiu com uma publicação, da autoria de Satoshi Nakamoto, intitulada de “Bitcoin: A Peer-to-Peer Electronic Cash System”, onde é descrito o modo de funcionamento de uma moeda digital, a *Bitcoin*, assente na tecnologia *Blockchain* (Härdle *et al.*, 2019). A 3 de janeiro de 2009, Nakamoto criou o chamado *Genesis Block*, o primeiro bloco a ser minerado e que tornou a *Bitcoin* real.

Em 2010, é registada a primeira transação comercial com *Bitcoins*, quando um programador chamado Laszlo Hanyecz compra duas pizzas por 10 mil *Bitcoins*, que na altura equivaleria a cerca de 30 dólares (Rose, 2015). Esta transação é considerada um momento significativo na história da *Bitcoin*, uma vez que prova que a criptomoeda poderia ser usada para transações do mundo real, fazendo assim despertar um interesse generalizado no potencial da tecnologia *Blockchain*.

Ao longo dos anos que se seguiram, o valor da *Bitcoin* começou a aumentar significativamente, atraindo cada vez mais a atenção dos investidores e tendo cada vez mais destaque nos meios de

comunicação. Em dezembro de 2017, registou-se o primeiro pico significativo no preço da *Bitcoin* alcançando o valor de 19 mil dólares, aproximadamente.

Durante esse período, começaram também a surgir novas criptomoedas, denominadas de *Altcoins* (combinação das palavras inglesas “Alternative coins”), sendo que a *Ethereum* é uma delas e, atualmente, a segunda maior em termos de capitalização de mercado.

A *Ethereum* é também uma rede *peer-to-peer*, composta por milhares de computadores de forma virtual e que qualquer pessoa pode usar para executar aplicações descentralizadas (*DApps*) ou até mesmo *smart contracts* dentro da própria *Blockchain* (Jani, 2017).

Os *smart contracts* são o ponto diferenciador da *Ethereum* quando comparada com a *Bitcoin* e com a maioria das criptomoedas que existem atualmente no mercado. Em conformidade com Kushwaha *et al.* (2022), os *smart contracts* ou contratos inteligentes são programas informáticos que executam tarefas previamente estabelecidas assentes em determinadas condições e aceites pelas partes que querem celebrar o contrato.

Em suma, o protocolo da *Ethereum* teve como primeira ideia a conceção de uma versão atualizada de uma criptomoeda que fornecesse recursos avançados, como por exemplo, o depósito dentro de uma *Blockchain*, limites nos levantamentos efetuados e a realização de *smart contracts* (Jani, 2017; Mónica, 2022).

O desenvolvimento tecnológico da *Ethereum* tem atualmente o apoio de mais de 100 empresas ao qual deram o nome de *Ethereum Enterprise Alliance* (EEA). Destas empresas podemos destacar: Acceture; BBVA; Cisco; Delloite; GoChain, no desenvolvimento de *DApps*; JP Morgan; Microsoft.

À data em que decorre a presente investigação², existem mais de 1.8 milhões de criptomoedas no mercado, que totalizam uma capitalização de mercado de mais de 1.05 triliões de dólares. De acordo com o *site CoinMarketCap*³, a *Bitcoin* e a *Ethereum* são as criptomoedas com maior capitalização de mercado, cada uma com cerca de 524.8 mil milhões de dólares e 186.6 mil milhões de dólares, respetivamente.

Com o surgimento diário de novas criptomoedas e o desaparecimento de outras, a especulação em torno destes investimentos também tem oscilado de dia para dia, provocando uma grande volatilidade nos preços. Este acontecimento tem levantado muitas questões sobre as quais muitos economistas e investigadores procuram resposta. Por exemplo, podemos prever o preço das criptomoedas? O que influencia o preço da *Bitcoin*? Existe alguma relação direta entre o preço da

Bitcoin e outros indicadores económicos? Apesar dos esforços registados, muitas dessas questões permanecem em aberto.

2.2. Preço das Criptomoedas e fatores que influenciam

Um marco notável na evolução das criptomoedas foi a criação de mercados especializados (*exchanges*) para o comércio e transação das mesmas. Deste modo, seria teoricamente estabelecida uma taxa de câmbio em tempo real, em que os investidores saberiam qual o valor atual da respetiva criptomoeda que pretendem transacionar em comparação com o valor do dólar. Estes *exchanges* criaram um serviço facilmente acessível e com a capacidade de transacionar criptomoedas em poucos segundos, bem como a disponibilidade de informação sobre preços em tempo real (Li & Wang, 2017).

O preço das criptomoedas desempenha um papel central na decisão dos investidores, pelo que a sua compreensão é fundamental para qualquer pessoa que deseja investir neste mercado. À medida que se tem verificado um crescimento exponencial em termos de notoriedade, volume transacionado, capitalização de mercado e número de criptomoedas oferecidas, maior tem sido a necessidade de tentar compreender os fatores que moldam os preços.

Deste modo, o comportamento do preço das criptomoedas e em particular o preço da *Bitcoin*, que por ser a primeira com sucesso e a que inspira maior certeza (dentro de um ecossistema de incerteza e volatilidade), tem despertado o interesse de muitos investigadores que, com base em diferentes modelos econométricos, procuram compreender ou prever esse comportamento.

Sovbetov (2018) publicou um artigo em que examina os fatores que influenciam os preços das cinco criptomoedas mais comuns, como a *Bitcoin*, *Ethereum*, *Dash*, *Litecoin* e *Monero*. Com o objetivo de construir algumas variáveis explicativas e que sejam referentes ao próprio mercado das criptomoedas, concebeu o Índice Crypto50 e do qual derivou métricas como capitalização total de mercado, volume de negociação e volatilidade. De modo a analisar a dinâmica de curto e longo prazo do preço das criptomoedas, o autor aplicou o modelo ARDL (*Autoregressive Distributed Lag*). Atendendo aos resultados do estudo e focando-se apenas na *Bitcoin*, chegou à conclusão de que o aumento da capitalização de mercado e do volume de negociação têm um impacto positivo no preço a curto e longo prazo, enquanto o fator de atratividade tem apenas a longo prazo. Por outro lado, a volatilidade das criptomoedas tem um impacto negativo no preço da *Bitcoin* a curto e a longo prazo.

Nos anos mais recentes à sua criação, a atividade de mercado das criptomoedas desempenhava um papel central na determinação do preço das mesmas, mas com os anos tem-se vindo a verificar uma diminuição da sua importância em prol de fatores económicos (Li & Wang, 2017).

Zhu *et al.* (2017) desenvolveram um estudo que visa analisar como é que fatores como o Índice de Preços no Consumidor (CPI), Índice do Dólar Americano (USDI), *Dow Jones Industrial Average* (DJIA), *Federal Funds Rate* (FFR) e Preço do Ouro (GP), influenciam o preço da *Bitcoin*. Para o efeito recorreram à utilização de um modelo VEC (*Vector Error Correction*) e concluíram que, a longo prazo, as variáveis CPI, DJIA, FFR e USDI influenciam negativamente o preço da *Bitcoin*. Quanto ao Preço do Ouro, concluem que este não tem qualquer tipo de influência a curto e a longo prazo.

Por outro lado, num estudo realizado por Moussa *et al.* (2021), no qual utilizam modelos de cointegração lineares e não lineares para estudar a dinâmica de ajuste de curto e longo prazo do preço da *Bitcoin*, verificaram que os preços logaritmizados do ouro afetam significativamente os preços logaritmizados da *Bitcoin*, no curto prazo.

Malladi *et al.* (2020) procederam à análise de séries temporais do retorno e volatilidade da *Bitcoin* e como estas se relacionam com Índice de Ações (índice S&P500, *MSCI World Index*, *MSCI Emerging Markets Index*), Preços do Ouro e Índices de Medo, como o VIX e o Índice de Incerteza da Política Económica dos EUA. Para o efeito, usaram os modelos ARMAX (*Autoregressive Moving Average with Extra Input*), GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*), VAR (*Vector Autoregressive*) e testes de causalidade de Granger. Com os resultados obtidos, concluíram que os retornos dos mercados de ações globais e do ouro não têm um efeito causal nos retornos da *Bitcoin*, sendo que os principais fatos que influenciam o preço da mesma são indicadores de medo e outros retornos de criptomoedas.

A comparação da *Bitcoin* ao ouro tem sido recorrente na literatura, com a discussão de que esta possa ser um substituto útil ao mesmo. O facto de ter a sua oferta previamente definida a um total de 21 milhões de moedas, ser descentralizada e distribuída e apontada igualmente como um produto de investimento são algumas das principais características de similaridade referenciadas. Todavia, a comparação não tem sido apenas com o ouro, pois a comparação com o dólar também tem sido discutida, principalmente por serem formas alternativas de moeda e reserva de valor.

Dyhrberg (2016) concluiu que a *Bitcoin* está entre uma moeda e uma *commodity*, isto é, tendo introduzido dois modelos GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*)

para analisar as semelhanças entre a *Bitcoin*, o ouro e o dólar e verificou que o comportamento do preço da *Bitcoin* apresenta semelhanças com o preço do dólar, reagindo significativamente à *Federal Funds Rate*, mas principalmente com o comportamento do preço do ouro, pois a capacidade de reação a boas e más notícias é muito semelhante.

A atenção do mundo acadêmico nem sempre se tem centrado em fatores macroeconômicos, principalmente no que respeita ao mercado de criptomoedas e comportamento de preços. Este é um mercado extremamente volátil e de resposta imediata, pelo que os investigadores têm dado também importância a fatores sociais e eventos que afetam o mundo.

Neste sentido, Guégan & Renault (2020) procuraram a resposta para a questão “Qual é o papel desempenhado pelo sentimento dos investidores no processo de dinâmica de preços?”. Como representação do sentimento dos investidores, recorreram à plataforma *StockTwits* para descarregar os dados relacionados com a *Bitcoin* no período de 1 de agosto de 2017 a 1 de dezembro de 2019 e construir uma métrica de sensibilidade. Utilizando Regressão OLS (*Ordinary Least Squares*) e Testes de Causalidade de Granger, os autores concluíram que o sentimento do investidor desempenha um papel importante na determinação do preço da *Bitcoin*. Para frequências de 15 minutos, verificaram que as mensagens enviadas no *StockTwits* durante o período $t - 1$ são estatisticamente significativas e têm um efeito positivo sobre o preço da *Bitcoin*.

Relacionada com a questão anterior, Sattarov *et al.* (2020) recorreram à rede social *Twitter*, para recolher *tweets* relacionados com a *Bitcoin*, e perceber se o preço da *Bitcoin* poderia ser estimado usando esses dados. Para o efeito, utilizaram o método VADER (*Valence Aware Dictionary and Sentiment Reasoner*) e encontraram evidências de que os *tweets* têm poder preditivo sobre o preço da *Bitcoin*. No estudo realizado, entre 12 de março e 12 de maio de 2018, observaram uma precisão de 62,48% ao fazer previsões.

A título de síntese, dir-se-á que diversas variáveis têm sido apontadas na literatura para explicar o comportamento do preço da *Bitcoin* ou prever o mesmo, tais como: variáveis macroeconômicas, a própria economia da *Bitcoin*, outras criptomoedas e variáveis de atratividade e sentimento.

CAPÍTULO 3

Metodologia

Nesta dissertação pretendemos estudar dados em formato de série temporal, caracterizados por um conjunto de observações de uma variável, dispostas sequencialmente no tempo, mostrando como os dados se ajustam ao longo do período. O objetivo deste estudo passa por aplicar e avaliar a capacidade preditiva do modelo ARIMA no que respeita ao preço da *Bitcoin* e, posteriormente, desenvolver um outro modelo da mesma família, incluindo diferentes variáveis exógenas referidas na revisão de literatura. A metodologia usada consta em vários passos sucessivos, tendo como base o método de Box-Jenkins (análise de séries temporais) combinado com o CRISP-DM (*Cross Industry Standard Process for Data Mining*).

Para criar e testar a capacidade de previsão dos nossos modelos, foi necessária uma forte componente computacional, desta forma optou-se pela utilização de um ambiente *Jupyter Notebook* na linguagem de programação *Python*⁴, uma vez que é uma ferramenta muito utilizada na exploração e transformação de dados, graças à vasta panóplia de bibliotecas que tem disponíveis. Alguns exemplos das bibliotecas utilizadas são: *numpy* (fornece funções matemáticas mais eficientes, tratando os dados em forma de matriz), *pandas* (utilizada para a transformação e análise de dados), *statsmodels* (contempla uma série de ferramentas para implantação e análises estatísticas), *matplotlib* (utilizado para a criação de diferentes visualizações gráficas) e *yfinance* (para a obtenção dos dados que serão estudados).

3.1. Estacionariedade

Por forma a que as inferências estatísticas dos modelos com séries temporais sejam válidas, o primeiro passo, depois de obter os dados, é validar a estacionariedade das mesmas, isto é, verificar se as séries temporais utilizadas seguem um processo estocástico com média e variância constantes ao longo do tempo e a covariância entre valores desfasados da série depende apenas do valor da “distância” temporal entre eles.

Para validar a natureza das variáveis nesta dissertação, foram realizados os seguintes testes de raiz unitária ou de estacionariedade: *Augmented Dickey Fuller* (ADF), *Phillips-Perron* (PP) e *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS).

No que respeita ao estudo da estacionariedade de uma série temporal, o teste ADF é dos mais utilizados. Este avalia se estamos na presença de uma raiz unitária ou não, indicando se a série é não-estacionária ou estacionária. Em notação algébrica, o teste ADF é descrito pela seguinte equação:

$$\Delta Y_t = (\rho - 1)y_{t-1} + \sum_{i=1}^k \Delta y_{t-i} + e_t$$

onde Y_t representa a série temporal no momento t , ρ e β são parâmetros e e_t é o termo de erro. Sendo um teste de significância estatística, envolve uma hipótese nula e alternativa, pelo que a hipótese nula pressupõe a existência de raiz unitária ($\rho = 1$), ou seja, a não estacionariedade da série.

O teste PP analisa, de igual modo, a estacionariedade de uma série temporal e, apesar de ter também como hipótese nula a existência de raiz unitária, tem as suas diferenças em relação ao teste ADF: ignora qualquer correlação e heteroscedasticidade nos resíduos.

Esta característica torna o teste mais robusto, pelo que quando aplicado à mesma série temporal é possível obter resultados contraditórios. Nestas situações surge o terceiro teste mencionado, o KPSS, baseado numa regressão linear. A expressão algébrica do teste divide a série em análise em três partes: passeio aleatório (r_t), tendência determinística (β_t) e erro estacionário (ε_t), conforme descrito na seguinte equação:

$$x_t = r_t + \beta_t + \varepsilon_t$$

A hipótese nula para este teste é de que a série é estacionária (ou seja, o passeio aleatório e/ou a tendência determinística anulam-se) e a hipótese alternativa é de que a série é não estacionária.

Neste sentido, tanto para o teste ADF como para o teste PP, quando o *p-value* é superior ao nível de significância, não se rejeita a hipótese nula e por esta razão a série é não-estacionária. Para o teste KPSS, não se rejeita igualmente a hipótese nula, mas a série é estacionária.

3.2. Causalidade de Granger

No que concerne ao estudo da causalidade entre as variáveis, decidiu aplicar-se o teste de Causalidade de Granger. Este teste mede a significância estatística da correlação entre o valor corrente de uma variável Y e os desfasamentos de outra variável X e, caso exista, diz-se que a variável X causa Granger a Y .

A hipótese nula para o teste é que os valores desfasados de X não explicam a variação em Y (ou seja, todos os coeficientes dos termos desfasados de X na equação de Y são nulos), contra a hipótese alternativa em que valores desfasados de X ajudam a explicar a variação em Y .

Assim, para p -values inferiores ao nível de significância, rejeitamos a hipótese nula, ou seja, existe causalidade de Granger, nomeadamente X causa Granger Y . Nos casos em que X causa Granger Y e, por outro lado, Y causa X , diz-se que estamos perante uma causalidade bidirecional.

3.3. Modelo ARIMA

Os modelos de séries temporais são uma tentativa de capturar características empiricamente relevantes dos dados observados para posteriormente projetá-las sobre as previsões. Uma classe importante de modelação de séries temporais é a família de modelos *Auto Regressive Integrated Moving Average* (ARIMA), que, quando implementada à uma série temporal e seguindo um número de passos (visualização, estudo da estacionariedade e sazonalidade, diferenciação, identificação das ordens/desfasamentos do modelo, estimação, testes de diagnóstico, validação e previsão), também é referida como metodologia de Box-Jenkins.

Existem diversos tipos de modelos ARIMA que diferem dependendo da sazonalidade e/ou uso de variáveis exógenas, mas cujo pressuposto base é o de estacionariedade da série. Caso não se verifique, deve proceder-se à transformação dos dados para torná-los estacionários, como por exemplo, aplicando as diferenças sucessivas ou diferenças logaritmizadas.

De modo geral, o modelo define-se por ARIMA (p, d, q) , onde p refere-se à ordem do processo autorregressivo incorporado no modelo, d refere-se à ordem de integração ou diferenciação da série e q refere-se à ordem do processo de média móvel incorporado no modelo. Mais especificamente, o modelo ARIMA é uma combinação entre um modelo autorregressivo, $AR(p)$, um modelo de média móvel, $MA(q)$, e a ordem de diferenciação da série, sendo representado pela seguinte equação:

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + \varepsilon_t$$

Utilizando o operador lag, a equação pode ser escrita de forma equivalente como sendo:

$$\Phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t$$

onde

$$\Phi(L) = 1 - a_1L - a_2L^2 - \dots - a_pL^p$$

$$\Theta(L) = 1 + \theta_1L + \theta_2L^2 + \dots + \theta_qL^q$$

Sendo os $\alpha_i, i = 1, \dots, p$ e $\theta_i, i = 1, \dots, q$ os coeficientes para estimar, e onde ε_t é um processo de ruído branco, isto é, processo de média nula, variância constante e ausência de autocorrelação, denotando-se por $\varepsilon_t \sim WN(0, \sigma^2)$.

Uma extensão deste modelo é o modelo *AutoRegressive Integrated Moving Average with Exogenous Variables* (ARIMAX), capaz de captar perturbações de variáveis exógenas e que possam ter impacto no modelo. Estas variáveis podem ser séries temporais adicionais ou variáveis independentes que não fazem parte da série temporal sujeita à modelação. O modelo ARIMAX pode ser traduzido pela seguinte equação:

$$y_t = c + a_1y_{t-1} + \dots + a_py_{t-p} - \theta_1e_{t-1} - \dots - \theta_qe_{t-q} + \beta_1X_{1,t} + \dots + \beta_mX_{m,t} + \varepsilon_t$$

onde β_1, \dots, β_m são os coeficientes das variáveis exógenas $X_{1,t}, \dots, X_{m,t}$.

A identificação do modelo ARIMA apropriado é conseguida, teoricamente, com recurso à transformação dos dados, caso seja necessário estacionarizar os mesmos, de modo a identificar o parâmetro d e à representação das funções de autocorrelação (ACF) e de autocorrelação parcial (PACF) das séries estacionárias de modo a identificar os parâmetros q e p , respetivamente.

Este método pode ser demorado e menos eficiente, tendo em conta os passos gerais necessários para a sua aplicação e comparativamente com a utilização direta da função `auto_arima()`, disponível na biblioteca `pmdarima` do `Python`. Também se tomou a decisão de completar os modelos, tanto em termos do número de *lags* como das variáveis exógenas a usar, tendo em consideração a compreensão do negócio.

A função está desenhada para automatizar o processo de seleção do modelo ARIMA que melhor se ajuste para um determinado conjunto de dados de séries temporais, através da pesquisa e análise automática de várias combinações dos parâmetros com base na minimizando do critério de informação de Akaike.

Portanto, quando confrontados com mais do que um modelo ARIMA, para a seleção do que melhor se ajusta aos dados, devemos recorrer a critérios de seleção. Os mais utilizados são os critérios baseados na maximização da função de verosimilhança, como é o ou, na minimização da

informação perdida, como é o caso do *Akaike Information Criterion* (AIC). A sua fórmula básica é dada por:

$$AIC = -2 \log(\mathcal{L}) + 2k$$

onde \mathcal{L} é o valor da função de máxima verosimilhança e k o número de parâmetros do modelo.

O modelo que melhor se ajusta aos dados será aquele cujo valor da estatística AIC é menor. No entanto, este critério em nada diz sobre a qualidade absoluta, isto é, se todos os modelos forem “pobres”, ele escolherá aquele que é melhor entre eles.

3.4. Previsão e Erro de Previsão

Uma vez selecionado, estimado e verificado o modelo, a próxima etapa é a previsão e posterior análise da capacidade preditiva do modelo.

Para ajuizar a qualidade dos modelos em termos da sua capacidade de previsão é necessário definir um conjunto de métricas de modo a avaliar o desempenho dos mesmos e a qualidade das previsões. As medidas tradicionalmente utilizadas para o efeito são baseadas na diferença entre os valores preditos e os valores observados, o chamado Erro de Previsão.

A medida mais comumente usada para calcular o erro é a métrica *Mean Absolute Percentage Error* (MAPE), uma vez que o seu resultado é apresentado em unidades percentuais e facilita o entendimento. A representação algébrica pode ser traduzida por:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

onde n é o número de previsões, \hat{y}_t são os valores preditos e y_t os valores observados. O melhor modelo de previsão será aquele para o qual o MAPE é menor.

Para além desta métrica, no presente estudo será também utilizada a métrica *Root mean squared error* (RMSE), que representa o desvio padrão dos resíduos, que por outras palavras, representa quão dispersos estão esses resíduos da linha de regressão. O RMSE é calculado como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

onde n é o número de previsões, \hat{y}_t são os valores preditos e y_t os valores observados. Quanto menor o seu valor, melhor o modelo para efeitos de previsão. Esta métrica é dada na mesma escala do que a variável original em estudo.

CAPÍTULO 4

Dados e análise de estatísticas descritivas dos dados

As séries temporais em análise são: preço da *Bitcoin* e *Ethereum*, variáveis representativas do mercado de criptomoedas (representam cerca de 70% da capitalização de mercado total), preço do ouro, preço do Petróleo, Índice acionista S&P500 e Índice de volatilidade do mercado de ações (VIX).

Os dados foram obtidos com recurso à base de dados do *Yahoo Finance*, com o período de análise compreendido entre 03 de junho de 2019 e 31 de maio de 2023. Ao conjunto dos dados, foram retirados apenas os valores de fecho de cada série.

Tendo em conta que os mercados de criptomoedas funcionam 24/7 e o das restantes variáveis se encontram encerrados aos fins-de-semanas e feriados, identificou-se a necessidade de utilizar uma periodicidade semanal dos dados. Posto isto, o *dataset* é composto por 209 semanas.

No que respeita à representação gráfica das séries temporais podemos *a priori* identificar algumas características: as séries Ouro, Petróleo e SP500 apresentam uma tendência não-linear crescente (Figura 1), a série VIX apresenta uma tendência não-linear decrescente (Figura 1), enquanto as séries *Bitcoin* e *Ethereum* não apresentam qualquer tendência padronizada (Figura 2). Esta última conclusão mostra claramente a instabilidade envolvente do universo das criptomoedas, com diferentes tendências decrescentes e crescentes durante o período.

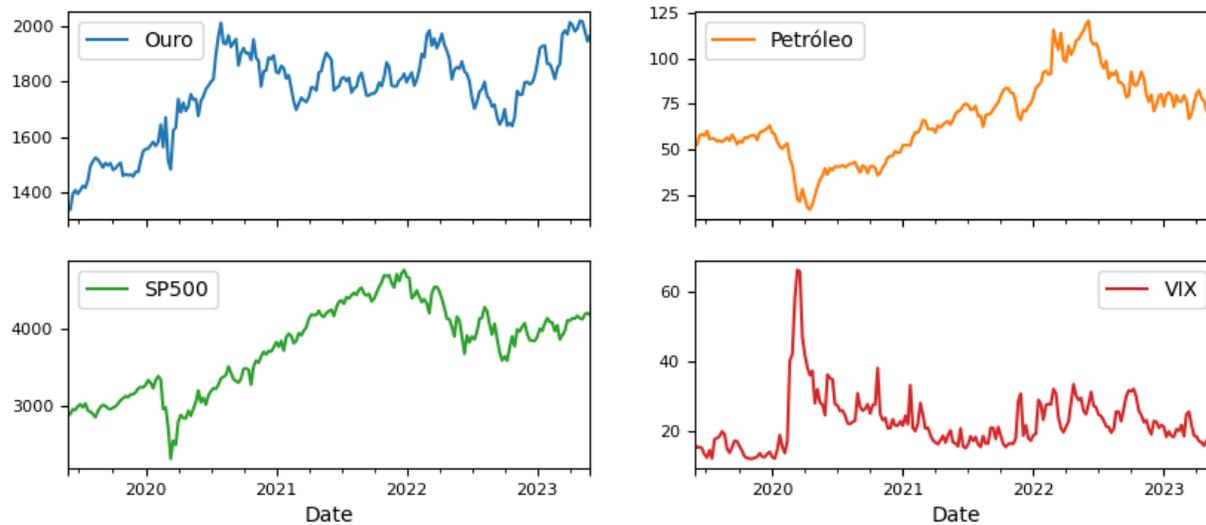


Figura 1. Representação gráfica das variáveis Ouro, Petróleo, SP500 e VIX.

Desta representação gráfica depreende-se, ainda, a existência de flutuações sazonais apenas na variável VIX, sendo esta aditiva.

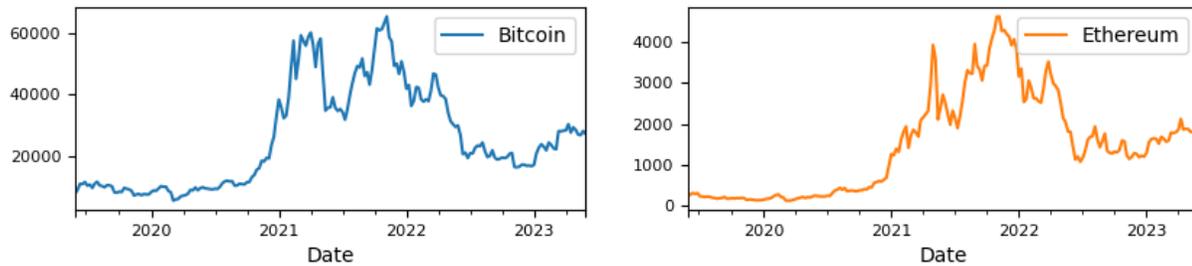


Figura 2. Representação gráfica do preço da Bitcoin e Ethereum.

A Tabela 1 apresenta algumas estatísticas descritivas dos dados. Complementarmente à representação gráfica, podemos verificar que os valores mínimos das variáveis Petróleo e SP500, 16.940\$ e 2304.920\$ respetivamente, e o valor máximo da variável VIX, 66.040\$, foram atingidos durante o período inicial da pandemia de COVID-19.

Por outro lado, podemos ainda verificar que a pandemia de COVID-19 não parece ter tido qualquer impacto significativo no preço da *Bitcoin* e *Ethereum*, pois veja-se que a partir de finais de 2020, os mesmos começaram a crescer de forma acentuada, atingindo os valores máximos de 65466.840\$ para a *Bitcoin*, em outubro de 2021, e 4626.359\$ para a *Ethereum*, em novembro de 2021.

As medidas que permitem interpretar a normalidade dos dados são a assimetria, que mede o grau de simetria da curva em relação à distribuição normal, e a curtose, que mede o grau de achatamento da curva em relação à curva normal.

Quanto à assimetria, dos valores apresentados na Tabela 1, podemos inferir que as variáveis *Bitcoin*, *Ethereum* e Petróleo são assimétricas positivas, sendo a última muito próxima de ser simétrica (0.182626). Por sua vez, as variáveis Ouro e SP500 são assimétricas negativas. De destacar o valor elevado de assimetria da variável VIX (2.170246), cujos dados apresentam alto grau de assimetria positiva.

Relativamente à curtose, todas as séries são consideradas platicúrticas, uma vez que o valor do coeficiente é inferior a zero, exceto a série VIX que é leptocúrtica, apresentando um valor de 7.981062. Da análise anteriormente feita, depreende-se a não normalidade de todas as séries em análise.

Variáveis	Média	Mediana	Máximo	Mínimo	Desvio-Padrão	Assimetria	Curtose
Bitcoin	25682.468	21027.295	65466.840	5392.315	16425.962	0.65536	-0.74397
Ethereum	1479.347	1338.636	4626.359	123.321	1223.372	0.67772	-0.48180
Ouro	1758.163	1785.300	2017.400	1340.100	158.808	-0.71579	-0.23469
Petróleo	66.693	66.260	120.670	16.940	22.086	0.18263	-0.28343
SP500	3760.994	3895.080	4766.180	2304.920	574.410	-0.27235	-1.02649
VIX	22.488	20.870	66.040	12.050	8.177	2.17025	7.98106

Tabela 1. Estatísticas descritivas das séries temporais.

De modo a analisar a correlação entre as variáveis foi efetuado um mapa de calor, representado na Figura 3. Dos valores obtidos podemos destacar a forte correlação positiva entre as variáveis *Bitcoin-Ethereum*, *Bitcoin-SP500* e *Ethereum-SP500*. As variáveis *Ouro* e *Petróleo* também possuem uma correlação positiva, embora não tão acentuada. De notar, ainda, que a variável *VIX* não possui qualquer correlação significativa com as restantes variáveis, variando entre os valores -0.13 e -0.27.



Figura 3. Mapa de calor de correlação das variáveis.

De seguida, para testar a direção da causalidade entre variáveis, recorreremos ao conceito de Causalidade de Granger.

Neste caso, foram usadas as variáveis com até quatro desfasamentos, sendo que os resultados apresentados (Tabela 2), tendo em conta que as variáveis em linha são as variáveis *target* (Y) e as

variáveis em coluna são as que podem causar a variável *target* (X), permitem-nos concluir que apenas o Ouro causa Granger *Bitcoin*, para um nível de significância de 15%.

As relações de causalidade que podemos também verificar são: *Bitcoin*, *Ethereum*, Ouro e VIX causam Granger SP500; *Bitcoin*, *Ethereum* e SP500 causam Granger Petróleo; *Bitcoin* causa Granger VIX e SP500 causa Granger Ouro, para um nível de significância de 15%. A única relação de causalidade bidirecional encontra-se entre as variáveis SP500 e Ouro.

Variáveis	Bitcoin	Ethereum	Ouro	Petróleo	SP500	VIX
Bitcoin	1.0000	0.1999	0.0850	0.3247	0.5520	0.1835
Ethereum	0.3465	1.0000	0.2444	0.4435	0.2629	0.3273
Ouro	0.2781	0.3140	1.0000	0.2629	0.0193	0.3058
Petróleo	0.0393	0.1076	0.0901	1.0000	0.0036	0.9393
SP500	0.0002	0.0004	0.0276	0.3922	1.0000	0.0001
VIX	0.0781	0.3216	0.5816	0.6822	0.2419	1.0000

Tabela 2. Teste de Causalidade de Granger.

Resultados e Discussão

5.1. Estacionariedade das séries

Com base na representação gráfica das séries temporais, é possível perceber se as mesmas são ou não estacionárias. No entanto, esta análise pode ou não ser conclusiva, pelo que o mais correto é utilizar testes de Raiz Unitária/Estacionariedade. A Tabela 3 resume os resultados obtidos para os mesmos.

Aplicado o teste ADF, os resultados indicam a não estacionariedade de todas as séries em níveis, uma vez que o p-value é superior a 0.05 para qualquer uma delas e por essa razão não rejeitamos a hipótese nula do teste (sendo a hipótese nula – existência de uma raiz unitária ou não-estacionariedade do processo estocástico). Os resultados do teste PP confirmam a não estacionariedade das séries em análise à exceção da série VIX, em que a conclusão é contraditória: rejeitamos a hipótese nula dado que o valor p-value é inferior a 0.05 e concluimos que estamos perante uma série estacionária. Procedendo à aplicação do teste KPSS, verificamos que as séries que foram classificadas como não estacionárias confirmam-se ser assim e quanto ao resultado contraditório da série VIX, segundo este teste a mesma é estacionária.

Resumindo, as séries *Bitcoin*, *Ethereum*, Ouro, Petróleo e SP500 são não estacionárias em níveis e a série VIX é estacionária em níveis.

Variáveis	ADF	PP	KPSS
Bitcoin	0.778153	0.760736	0.010000
Ethereum	0.770449	0.819519	0.010000
Ouro	0.198256	0.222496	0.010000
Petróleo	0.814261	0.691066	0.010000
SP500	0.693009	0.77277	0.010000
VIX	0.141186	0.029999	0.100000

Tabela 3. P-values dos testes de estacionariedade para as séries temporais em níveis.

Posto isto e para poder ser aplicado qualquer modelo econométrico, as séries não estacionárias terão de ser estacionarizadas, neste caso, através do cálculo das primeiras diferenças logaritmizadas (é de notar que se perde a primeira observação de cada série).

Para verificar a estacionariedade das séries é necessária a nova realização dos testes anteriormente analisados (Tabela 3), daí concluímos que todas as séries em causa são estacionárias na primeira diferença, dizendo-se por isso séries integradas de primeira ordem, $I(1)$.

Variáveis	ADF	PP	KPSS
Bitcoin	1.248357e-23	7.464520e-24	0.100000
Ethereum	4.248017e-24	5.386929e-24	0.100000
Ouro	3.006948e-14	3.069514e-29	0.10000
Petróleo	3.975535e-11	1.282147e-22	0.1000
SP500	3.558111e-28	8.880287e-29	0.100000

Tabela 2. P-values dos testes de estacionariedade para as primeiras diferenças logaritmizadas das séries temporais.

5.2. Estimação do modelo e previsão

Uma vez conseguida a estacionariedade das séries *Bitcoin*, *Ethereum*, *Ouro*, *Petróleo* e *SP500*, através da aplicação das primeiras diferenças logaritmizadas, estamos agora na posse de aplicar modelos econométricos e encontrar aquele que melhor se ajusta aos dados em estudo.

Neste sentido, optou-se por estudar a adequabilidade do modelo ARIMA, abreviação de *Auto Regressive Integrated Moving Average*, e verificar como este se ajusta aos dados da série temporal *Bitcoin*. A escolha deste modelo prende-se com o facto de este ser adequado para previsões de curto prazo.

Antes de prosseguirmos com o estudo, procedeu-se à divisão do conjunto de dados em treino e em teste da seguinte forma: o conjunto de treino é composto pelos registos das semanas de 03 de junho de 2019 a 26 de março de 2023 e o conjunto de teste é composto pelos registos das semanas de 27 de março de 2023 a 28 de maio de 2023. Assim, o conjunto de treino é composto por 199 semanas e o de teste por 10 semanas. Para os resultados obtidos dos modelos consideramos o nível de significância de 5%.

A escolha da janela de previsão ($k = 10$ semanas) não foi aleatória, sendo feita uma simulação da variação da performance de previsão em relação ao comprimento da janela de previsão. Observou-se que para menos de 10 semanas as previsões são feitas por excesso, enquanto para uma janela superior a 10 semanas as previsões são feitas por redução comparando com o valor real da *Bitcoin* (Anexo A). O desempenho do modelo ARIMA para a previsão dos preços da criptomoeda

Bitcoin é ilustrada na Figura 4. Observa-se ainda, o aumento da proximidade da série predita em relação a séries dos preços reais do *Bitcoin*, quando adicionamos mais momentos históricos na parte autorregressiva, sendo o valor ótimo 4 semanas.

Em primeiro lugar, foi aplicada a função *auto_arima*, que analisa múltiplas hipóteses do modelo para identificar os parâmetros q , d e p ótimos, sem a atribuição de qualquer tipo de

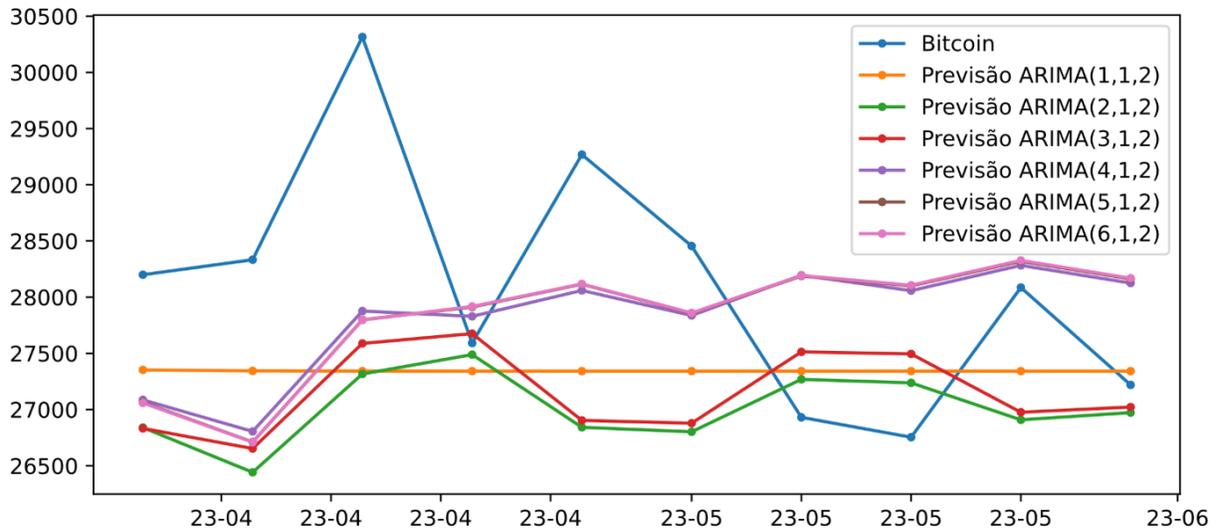


Figura 4. Previsão de 10 semanas com o modelo ARIMA, para vários lags na parte AR.

parâmetros, ou seja, foi apenas aplicada a função ao conjunto de dados em estudo (treino) da *Bitcoin*, com os restantes valores a *default*.

Para este cenário, obtivemos que o modelo ótimo é o ARIMA(2,1,2). Analisado o teste de *Ljung-Box* que sugere um $p\text{-value} < 0.05$, podemos afirmar que estamos na presença de resíduos de ruído branco, ou seja, não há autocorrelação nos mesmos, sendo, portanto, independentes e, assim, o modelo é apto para fazer previsões (Anexo A). Apesar do modelo ser válido e o mais correto de ponto de vista econométrico, a sua capacidade de previsão foi ultrapassada por outros modelos, como pode ser comprovado pela Figura 4 acima.

De seguida, com o objetivo de introduzir as variáveis *Ethereum*, Ouro, Petróleo, SP500 e VIX como variáveis exógenas, procedeu-se à determinação de um novo modelo, neste caso tratando-se de um modelo ARIMAX (*AutoRegressive Integrated Moving Average with Exogenous Variables*).

Para o efeito, inicialmente foram feitas várias combinações com as variáveis exógenas a serem consideradas na determinação do modelo. Tendo em conta cada uma das variáveis, assim como conjuntos de duas, três, quatro e cinco variáveis, obtivemos 31 combinações diferentes.

Posto isto, a função *auto_arima* foi utilizada novamente, aplicando-a ao conjunto de dados em estudo tantas vezes quantas as combinações identificadas, especificando a combinação a considerar como variáveis exógenas para cada determinação do modelo. Para qualquer uma das combinações identificadas o modelo ótimo obtido foi sempre o ARIMAX(0,1,1), pelo que optamos por verificar de forma menos “automatizada” o desempenho dos modelos ARIMAX.

Concluimos que a biblioteca de aprendizagem automatizada não consegue captar o efeito da adição de variáveis exógenas no modelo ARIMA. Sendo assim, voltamos a estimar vários modelos, usando a função *SARIMAX* da biblioteca *statsmodels* com o conhecimento adquirido pela análise automatizada e o juízo de valor baseado na compreensão do negócio.

Como modelo base usámos o ARIMA(4,1,2) e adicionámos as variáveis exógenas a este modelo. De entre as várias experiências destaca-se a inclusão da variável exógena Ouro, sendo a única causal pelo teste de Granger. Como efeito a qualidade da previsão aumentou de forma bastante significativa.

O outro caso a salientar, consta na inclusão da variável *Ethereum* no ARIMA(4,1,2), sendo esta escolha justificada pelo facto de esta ser igualmente uma criptomoeda. Os valores preditos aproximaram-se ainda mais dos valores reais da variável alvo.

Finalmente, a inclusão da variável SP500, do Petróleo e do VIX, ficam aquém dos resultados obtidos pela inclusão das outras duas variáveis exógenas. Quaisquer outras combinações de variáveis exógenas resultam em previsões menos qualitativas do que do modelo base.

Inferido que todos os modelos se encontram aptos para fazer previsões, procedeu-se à representação gráfica dos valores estimados por cada modelo e dos valores observados (Figura 5). Através da análise gráfica, verificámos que a representação dos valores preditos pelo modelo ARIMAX(4,1,2) com a variável exógena *Ethereum* se situam o mais próximo da representação dos valores reais da *Bitcoin*, pelo que à partida nos faz supor que o mesmo é melhor para fazer previsões. Ainda se observa que a os valores preditos pelos modelos ARIMA(4,1,2) e ARIMAX(4,1,2) com variável exógena Ouro, são muito semelhantes e abaixo dos valores reais dos preços de fecho da *Bitcoin*.

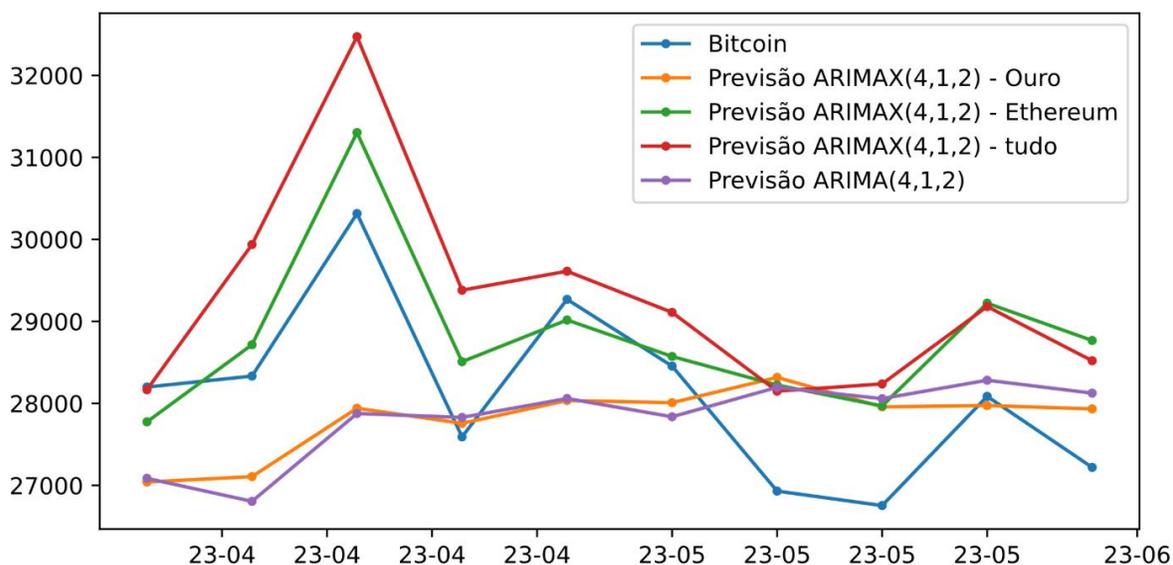


Figura 5. Representação gráfica do conjunto de teste e valores preditos para o Modelos ARIMA e ARIMAX(4,1,2).

Na Figura 5 ainda podemos observar a previsão quando no modelo ARIMAX usamos todas as variáveis em estudo – *Ethereum*, Ouro, SP500, Petróleo e VIX - como exógenas. A linha vermelha que corresponde a este modelo, mostra uma previsão por excesso, apesar de ter um padrão bastante semelhante com a da *Bitcoin*.

Para efeitos de confirmação algébrica e a fim de determinar o modelo que melhor se ajusta aos dados, a análise gráfica foi complementada com a comparação dos valores do critério de informação de Akaike (AIC) e do erro de previsão de cada um, medido pelas métricas *Mean Absolute Percentage Error* (MAPE) e *Root mean squared error* (RMSE).

Critérios	ARIMA(4,1,2)	ARIMAX(4,1,2) todas as variáveis	ARIMAX(4,1,2) Ouro	ARIMAX(4,1,2) <i>Ethereum</i>
AIC	3667.57	3533.45	3667.62	3527.01
MAPE	3.81%	4.15%	3.53%	2.97%
RMSE	1248.21	1324.82	1188.26	951.48

Tabela5. Valores do critério de informação AIC e métricas de erro MAPE e RMSE dos modelos ARIMA(4,1,2) e ARIMAX(4,1,1).

Face aos resultados representados na Tabela 5 acima, verificamos que o modelo com inferior AIC é o ARIMAX(4,1,2) – *Ethereum* com o valor de 3527.01, pelo que indicia ser o modelo que melhor se ajusta aos dados.

Recorrendo às métricas de medição de erro, podemos verificar que a conclusão imediatamente antes proferida é confirmada. Analisando o valor do MAPE, os valores preditos pelo ARIMAX(4,1,2) – *Ethereum* estão cerca de 2.97% desfasados do valor real. Adicionalmente, segundo a análise do valor do RMSE, a previsão com o este modelo está com um desfasamento de cerca de 951 dólares, em média, do preço real da *Bitcoin*.

Concluindo, tanto pelo erro absoluto médio em percentagem como pela raiz do erro quadrático médio, o modelo ARIMAX(4,1,2) – *Ethereum* tem uma capacidade de previsão melhor.

O modelo ARIMAX(4,1,2) – com todas as variáveis exógenas, apesar de mostrar o segundo menor valor do AIC, apresenta os maiores valores nas métricas de performance (MAPE e RMSE), indicando uma capacidade de previsão mais fraca, em comparação com todos os modelos apresentados.

Conclui-se que, para efeitos de previsão, a escolha do comprimento da janela de previsão é relevante (tendo a ser influenciada pelas componentes determinísticas da série temporal), também como, neste caso, a adição de variáveis exógenas causais – contribuem para o aumento da qualidade de previsão. Igualmente, a adição de uma variável não-causal (à Granger), mas pertencendo à mesma classe (criptomoeda), pode trazer benefícios em termos de previsão.

Do exposto aqui, conclui-se também que os modelos que melhor se ajustam aos dados podem não ter a melhor performance de previsão, sendo a compreensão do negócio um aspeto fundamental na modelação e na tomada de decisão.

Conclusão e recomendações

A *Bitcoin* surge numa época de grande desconfiança nos sistemas financeiros tradicionais e nas instituições bancárias, dada a Crise Financeira de 2008. Esta crise despertou preocupações no que respeita ao controlo centralizado do dinheiro e a necessidade de alternativas descentralizadas, iniciando uma revolução no sistema financeiro e tecnológico.

A escolha da temática foi despertada pelo crescente protagonismo que a *Bitcoin* tem vindo a ganhar como uma alternativa inovadora ao sistema financeiro tradicional. A exemplo disso, em Portugal, temos a existência de empresas cujo método de pagamento já pode ser em criptomoedas e, por outro lado, a existência de ATM (*Automatic Teller Machine*) para levantamento de criptomoedas. Outro fator de interesse prende-se com o estudo do comportamento do preço da *Bitcoin* que revela uma grande volatilidade e, em comparação com o preço do ouro, onde registou um crescimento exponencial num curto espaço de tempo.

Nesta dissertação analisámos a série temporal do preço da *Bitcoin* no horizonte de 03 de junho de 2019 a 31 de maio de 2023 e confirmámos a instabilidade do mesmo. Posteriormente, com o objetivo de analisar o modelo ARIMA que melhor se ajusta aos dados para efeitos de previsão sobre o preço da *Bitcoin*, realizou-se o estudo empírico inicialmente de forma automatizada obtendo os modelos ARIMA(2,1,2) e ARIMAX(0,1,1).

Perante este resultado insatisfatório, optou-se por analisar o desempenho dos modelos ARIMA de forma menos “automatizada”. Daqui aferimos que esta técnica é mais promissora para efeitos de previsão desta série.

Dos resultados empíricos obtidos por este método o modelo ARIMA(4,1,2) é o que apresenta uma melhor capacidade de previsão, razão pela qual foi a base de modelação para o modelo ARIMAX, com inclusão das variáveis exógenas. Exposto isto e atendendo à correlação e causalidade de Granger, concluímos ainda que a variável representativa da *Ethereum* é a que tem maior impacto na *Bitcoin*, seguindo-se a variável Ouro. Por outro lado, apesar de estudos que comprovam a relação da *Bitcoin* com as variáveis SP500, Petróleo e VIX, o presente estudo vem mostrar que as mesmas não são significativas à luz da metodologia aplicada.

As conclusões desta investigação proporcionam contributos tanto para a Academia, no sentido em que vem adicionar um novo caso de aplicação do modelo ARIMA e ARIMAX com uma

combinação de variáveis exógenas distinta. Por outro lado, é também do interesse dos investidores no mercado das criptomoedas utilizar as ferramentas aqui explicadas para tentar prever o preço da *Bitcoin* e adaptar as decisões de investimento em conformidade.

Sendo esta uma temática recente e ainda pouco abordada na literatura académica é fulcral continuar a desenvolver estudos sobre a mesma. No que concerne a futuras investigações e utilizando a metodologia aqui apresentada é possível adaptá-la a outras criptomoedas relevantes para o estudo da previsão. Por outro lado, a alteração das variáveis exógenas ou incorporação de outras de diferentes naturezas poderá permitir outras conclusões.

Por fim, importa-me referir que as criptomoedas não devem ser vistas como uma ameaça à indústria tradicional financeira. Pelo contrário, o objetivo é trazer competição a uma indústria estagnada e acomodada. Através desta competição idealizam-se novos produtos que contribuem para o avanço do setor económico e financeiro.

Referências Bibliográficas

- Antena 3. (2022, janeiro 23). Diogo Mónica | A Minha Geração com Diana Duarte | Antena 3 [Video]. YouTube. <https://www.youtube.com/watch?v=OrgyL7b7FAY>
- Dai, W. (1998) B-Money. <http://www.weidai.com/bmoney.txt>
- Doran, M. (2015). A Forensic Look at Bitcoin Cryptocurrency.
- Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar – A GARCH volatility analysis. *Finance Research Letters*, 16, 85–92. <https://doi.org/10.1016/j.frl.2015.10.008>
- European Central Bank. (2012). Virtual currency schemes. European Central Bank. <https://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemes201210en.pdf>
- Farell, R. (2015, May 1). An Analysis of the Cryptocurrency Industry. <https://repository.upenn.edu/handle/20.500.14332/49177>
- Guégan, D., & Renault, T. (2020). Does investor sentiment on social media provide robust information for Bitcoin returns predictability? *Finance Research Letters*, 101494. <https://doi.org/10.1016/j.frl.2020.101494>
- Gunay, S. (2019). Impact of public information arrivals on cryptocurrency market: A case of twitter posts on ripple. *East Asian Economic Review*, 23(2), 149–168. <https://doi.org/10.11644/kiep.eaer.2019.23.2.359>
- Jani, S. (2017). An overview of Ethereum & its comparison with bitcoin. *International Journal of Scientific & Engineering Research*, 10(8), 1-6.
- Kushwaha, S. S., Joshi, S., Singh, D., Kaur, M., & Lee, H.-N. (2022). Ethereum Smart Contract Analysis Tools: A Systematic Review. *IEEE Access*, 10, 57037– 57062. <https://doi.org/10.1109/access.2022.3169902>
- Li, X., Wang, C. (2017). The Technology and Economic Determinants of Cryptocurrency Exchange Rates: The Case of Bitcoin. *Decision support systems*, 95, 49-60. <https://doi.org/10.1016/j.dss.2016.12.001>
- Maese, V. A., Avery, A. W., Naftalis, B. A., Wink, S. P., & Valdez, Y. D. (2016). Cryptocurrency: A primer. *Banking Lj*, 133, 468.

- Malladi, R. K., & Dheeriya, P. L. (2020). Time series analysis of Cryptocurrency returns and volatilities. *Journal of Economics and Finance*, 45(1), 75–94. <https://doi.org/10.1007/s12197-020-09526-4>
- Mattila, J. (2016). The blockchain phenomenon. *Berkeley Roundtable of the International Economy*, 16.
- Moussa, W., Mgadmi, N., Béjaoui, A., & Regaieg, R. (2021). Exploring the dynamic relationship between Bitcoin and commodities: New insights through STECM model. *Resources Policy*, 74, 102416. <https://doi.org/10.1016/j.resourpol.2021.102416>
- Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. <https://bitcoin.org/bitcoin.pdf>
- Pacheco, A. (2021). *Bitcoin (7th Edition)*. Editora Self.
- Rodrigues, V. (2019). *A tecnologia blockchain: criptomoedas e tokens de investimento – desafios jurídico-fiscais [Dissertação de mestrado não publicada]*. Faculdade de Direito da Universidade de Coimbra.
- Rose, C. (2015). The Evolution of Digital Currencies: Bitcoin, A Cryptocurrency Causing A Monetary Revolution. *International Business & Economics Research Journal (IBER)*, 14(4), 617-622. <https://doi.org/10.19030/iber.v14i4.9353>
- Sattarov, O., Jeon, H. S., Oh, R., & Lee, J. D. (2020). Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis. *2020 International Conference on Information Science and Communications Technologies (ICISCT)*. <https://doi.org/10.1109/icisct50599.2020.9351527>
- Sovbetov, Y. (2018). Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litecoin, and Monero *Journal of Economics and Financial Analysis*. *Journal of Economics and Financial Analysis*, 2(2), 1–27. <https://doi.org/10.1991/jefa.v2i2.a16>
- Zhu, Y., Dickinson, D., & Li, J. (2017). Analysis on the influence factors of Bitcoin's price based on VEC model. *Financial Innovation*, 3(1). <https://doi.org/10.1186/s40854-017-0054-0>

ANEXO A

A. 1. MODELOS ARIMA: SELEÇÃO

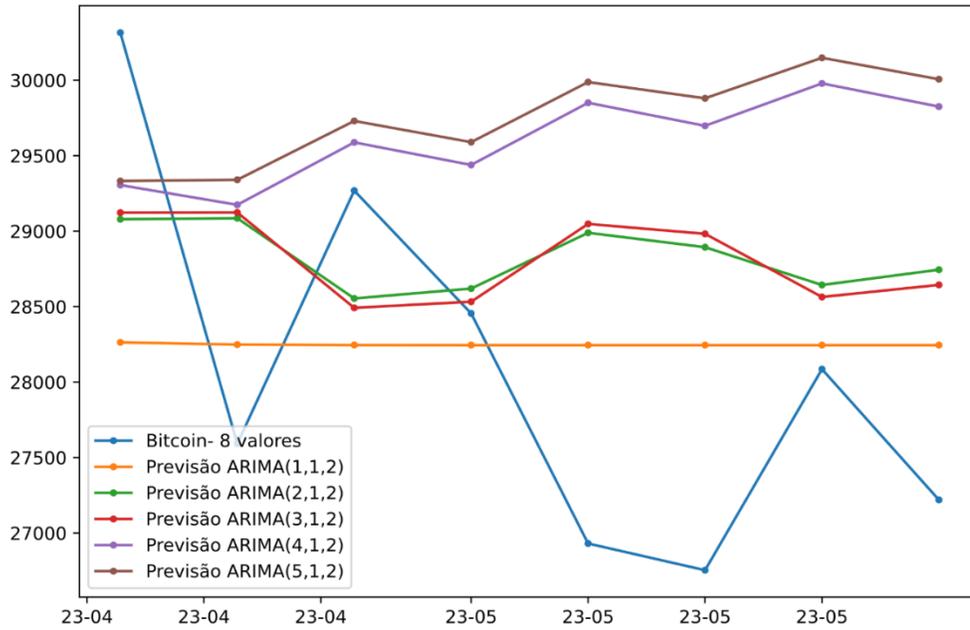


Figura A.1. Previsão de 8 semanas com o modelo ARIMA, para vários lags na parte AR.

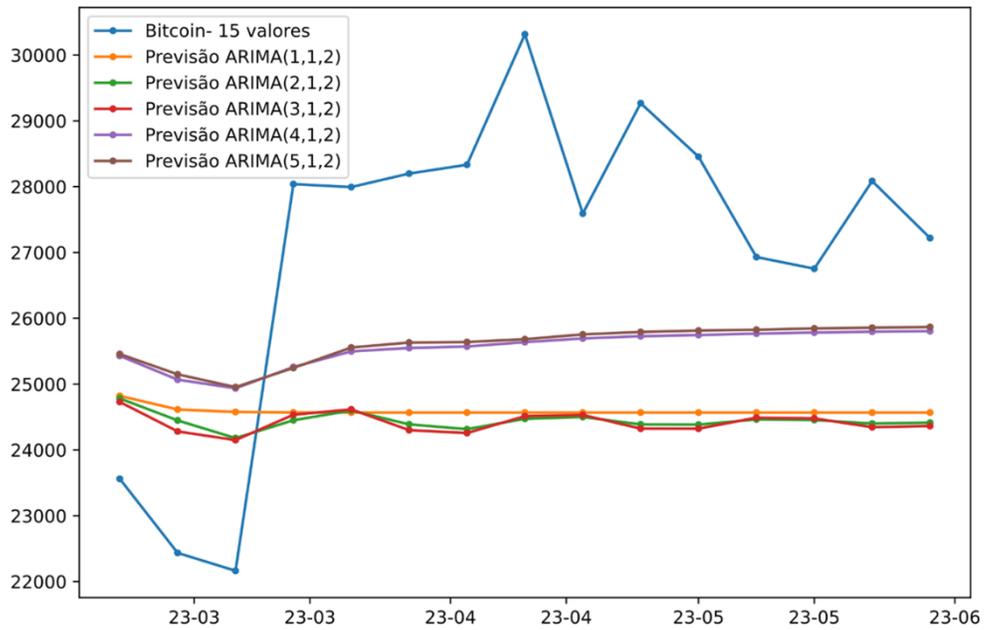


Figura A.2. Previsão de 15 semanas com o modelo ARIMA, para vários lags na parte AR.

ANEXO B

B. 1. MODELOS ARIMA: ESTIMAÇÃO POR VIA FUNÇÃO *AUTO_ARIMA()*

SARIMAX Results						
Dep. Variable:	y	No. Observations:	199			
Model:	SARIMAX(2, 1, 2)	Log Likelihood	-1864.473			
Date:	Sun, 29 Oct 2023	AIC	3738.945			
Time:	13:17:59	BIC	3755.387			
Sample:	06-03-2019	HQIC	3745.600			
	- 03-20-2023					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.1464	0.153	-0.960	0.337	-0.445	0.153
ar.L2	-0.7176	0.121	-5.938	0.000	-0.954	-0.481
ma.L1	0.2944	0.164	1.792	0.073	-0.028	0.616
ma.L2	0.6285	0.147	4.288	0.000	0.341	0.916
sigma2	9.328e+06	6.39e+05	14.608	0.000	8.08e+06	1.06e+07
Ljung-Box (L1) (Q):		0.04	Jarque-Bera (JB):	87.83		
Prob(Q):		0.85	Prob(JB):	0.00		
Heteroskedasticity (H):		8.51	Skew:	-0.39		
Prob(H) (two-sided):		0.00	Kurtosis:	6.17		

Figura B.1. Modelo ARIMA estimado por vias da função *auto_arma()*: série Bitcoin.

B. 2. MODELOS ARIMA: ESTIMAÇÃO POR VIA FUNÇÃO *SARIMAX*

SARIMAX Results						
Dep. Variable:	Bitcoin	No. Observations:	199			
Model:	SARIMAX(4, 1, 2)	Log Likelihood	-1826.785			
Date:	Sun, 29 Oct 2023	AIC	3667.570			
Time:	13:41:22	BIC	3690.445			
Sample:	06-03-2019	HQIC	3676.832			
	- 03-20-2023					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.0356	0.230	-0.155	0.877	-0.487	0.415
ar.L2	0.3848	0.220	1.748	0.081	-0.047	0.816
ar.L3	-0.1206	0.044	-2.714	0.007	-0.208	-0.034
ar.L4	0.2392	0.067	3.548	0.000	0.107	0.371
ma.L1	0.2077	0.239	0.870	0.384	-0.260	0.676
ma.L2	-0.5043	0.261	-1.933	0.053	-1.015	0.007
sigma2	9.068e+06	6.09e+05	14.903	0.000	7.88e+06	1.03e+07
Ljung-Box (L1) (Q):		0.01	Jarque-Bera (JB):	102.14		
Prob(Q):		0.92	Prob(JB):	0.00		
Heteroskedasticity (H):		9.29	Skew:	-0.56		
Prob(H) (two-sided):		0.00	Kurtosis:	6.37		

Figura B.2. Modelo ARIMA(4,1,2): série Bitcoin.

SARIMAX Results						
Dep. Variable:	Bitcoin	No. Observations:	199			
Model:	SARIMAX(4, 1, 2)	Log Likelihood	-1825.814			
Date:	Sun, 29 Oct 2023	AIC	3667.627			
Time:	13:55:16	BIC	3693.770			
Sample:	06-03-2019	HQIC	3678.213			
	- 03-20-2023					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
Ouro	6.1116	6.294	0.971	0.332	-6.225	18.448
ar.L1	-0.0481	0.215	-0.223	0.823	-0.470	0.374
ar.L2	0.4443	0.203	2.191	0.028	0.047	0.842
ar.L3	-0.1296	0.045	-2.871	0.004	-0.218	-0.041
ar.L4	0.2409	0.065	3.679	0.000	0.113	0.369
ma.L1	0.2277	0.222	1.024	0.306	-0.208	0.664
ma.L2	-0.5627	0.247	-2.277	0.023	-1.047	-0.078
sigma2	8.997e+06	1.28e-06	7.02e+12	0.000	9e+06	9e+06
Ljung-Box (L1) (Q):		0.01	Jarque-Bera (JB):	107.11		
Prob(Q):		0.91	Prob(JB):	0.00		
Heteroskedasticity (H):		10.41	Skew:	-0.54		
Prob(H) (two-sided):		0.00	Kurtosis:	6.47		

Figura B.3. Modelo ARIMAX(4,1,2) – Ouro.

SARIMAX Results						
Dep. Variable:	Bitcoin	No. Observations:	199			
Model:	SARIMAX(4, 1, 2)	Log Likelihood	-1755.460			
Date:	Sun, 29 Oct 2023	AIC	3526.921			
Time:	13:59:31	BIC	3553.063			
Sample:	06-03-2019	HQIC	3537.507			
	- 03-20-2023					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
Ethereum	9.9530	0.461	21.602	0.000	9.050	10.856
ar.L1	-1.6264	0.049	-33.278	0.000	-1.722	-1.531
ar.L2	-0.7670	0.080	-9.551	0.000	-0.924	-0.610
ar.L3	0.0352	0.082	0.431	0.666	-0.125	0.195
ar.L4	0.0345	0.058	0.590	0.555	-0.080	0.149
ma.L1	1.7995	0.039	46.040	0.000	1.723	1.876
ma.L2	0.9556	0.043	22.281	0.000	0.872	1.040
sigma2	4.411e+06	2.99e-08	1.48e+14	0.000	4.41e+06	4.41e+06
Ljung-Box (L1) (Q):		0.03	Jarque-Bera (JB):	148.85		
Prob(Q):		0.87	Prob(JB):	0.00		
Heteroskedasticity (H):		2.85	Skew:	-0.53		
Prob(H) (two-sided):		0.00	Kurtosis:	7.16		

Figura B.4. Modelo ARIMA(4,1,2) – Ethereum.

```

=====
SARIMAX Results
=====
Dep. Variable:          Bitcoin    No. Observations:          199
Model:                 SARIMAX(4, 1, 2)  Log Likelihood             -1757.507
Date:                 Sun, 29 Oct 2023  AIC                        3539.013
Time:                 14:01:25      BIC                        3578.227
Sample:               06-03-2019      HQIC                       3554.892
                   - 03-20-2023
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025      0.975]
-----
Ethereum      9.1116      0.520     17.520     0.000      8.092     10.131
Ouro          0.5268      6.326      0.083     0.934     -11.872     12.925
Petr3leo     49.7032     62.211      0.799     0.424     -72.229     171.635
SP500        0.7782      3.218      0.242     0.809     -5.529      7.086
VIX         -47.0026     80.709     -0.582     0.560    -205.188     111.183
ar.L1        0.0943      0.077      1.230     0.219     -0.056      0.245
ar.L2        0.8807      0.104      8.496     0.000      0.678      1.084
ar.L3       -0.1172      0.044     -2.642     0.008     -0.204     -0.030
ar.L4        0.0651      0.053      1.219     0.223     -0.040      0.170
ma.L1        0.0287      0.280      0.102     0.918     -0.520      0.577
ma.L2       -1.0248      0.110     -9.289     0.000     -1.241     -0.809
sigma2       4.243e+06    6.76e-05    6.27e+10    0.000    4.24e+06    4.24e+06
=====
Ljung-Box (L1) (Q):          0.09  Jarque-Bera (JB):          196.68
Prob(Q):                    0.76  Prob(JB):                   0.00
Heteroskedasticity (H):     2.96  Skew:                       -0.05
Prob(H) (two-sided):        0.00  Kurtosis:                   7.93
=====

```

Figura B.3. Modelo ARIMAX(4,1,2) – Ethereum, Ouro, SP500 e VIX.

ANEXO C

C. 1. MODELOS ARIMA: PREVISÃO

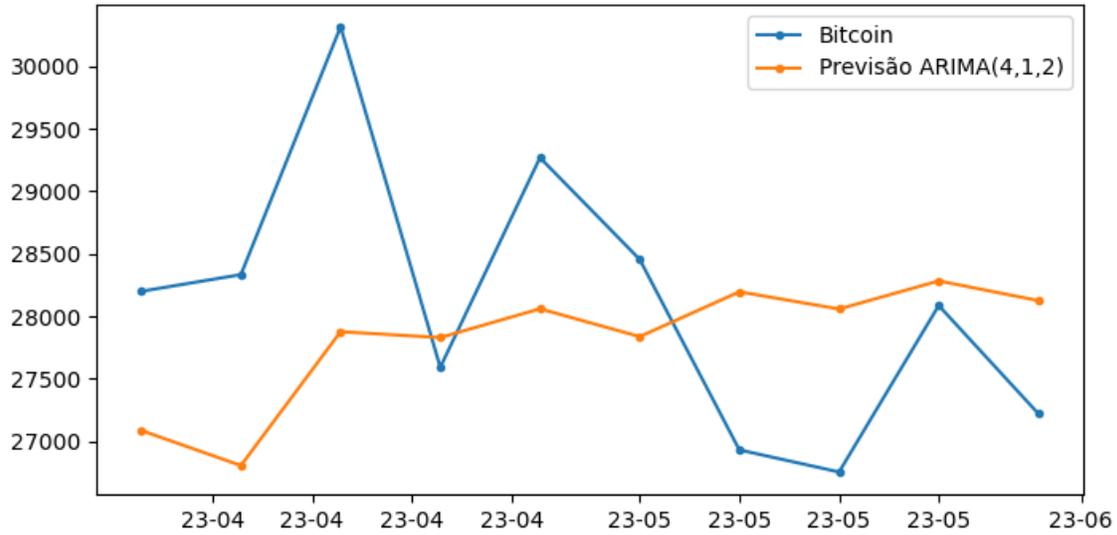


Figura C.1. Representação gráfica do conjunto de teste e valores preditos para o Modelos ARIMA 4,1,2).

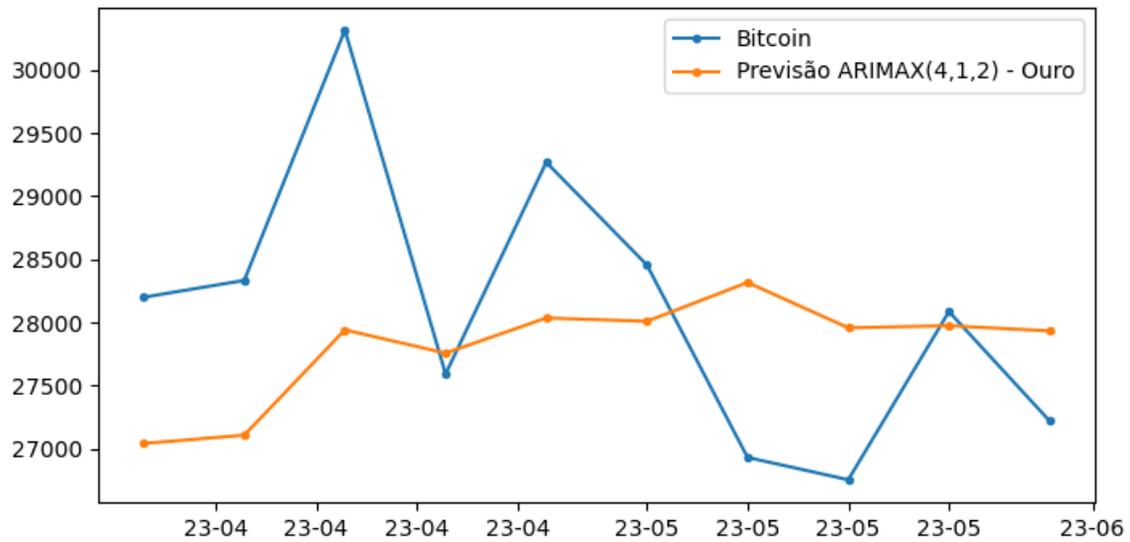


Figura C.2. Representação gráfica do conjunto de teste e valores preditos para o Modelos ARIMAX 4,1,2) - Ouro.

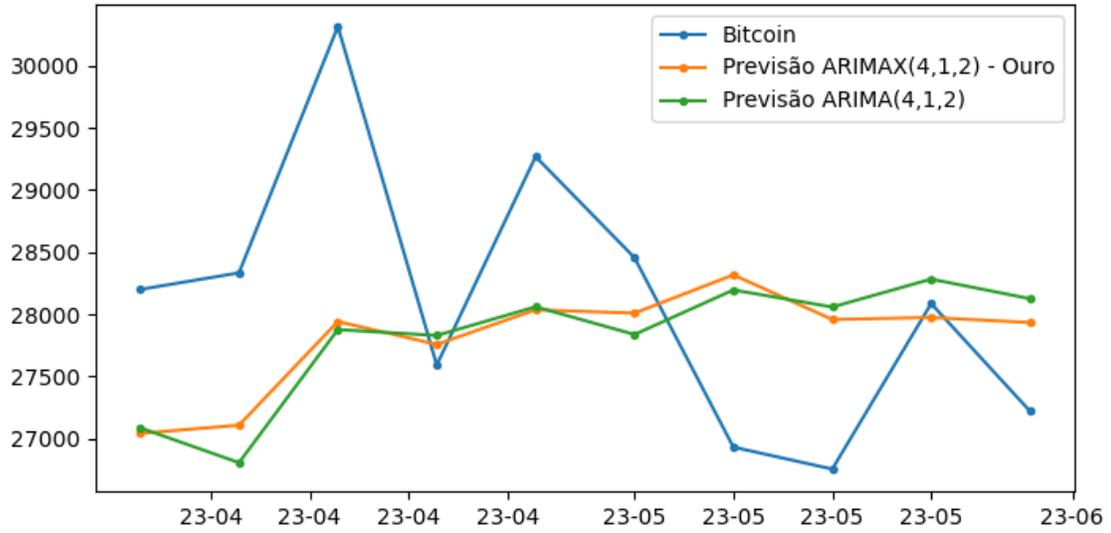


Figura C.3. Representação gráfica do conjunto de teste e valores preditos para o Modelos ARIMA 4,1,2) e ARIMAX()- Ouro.

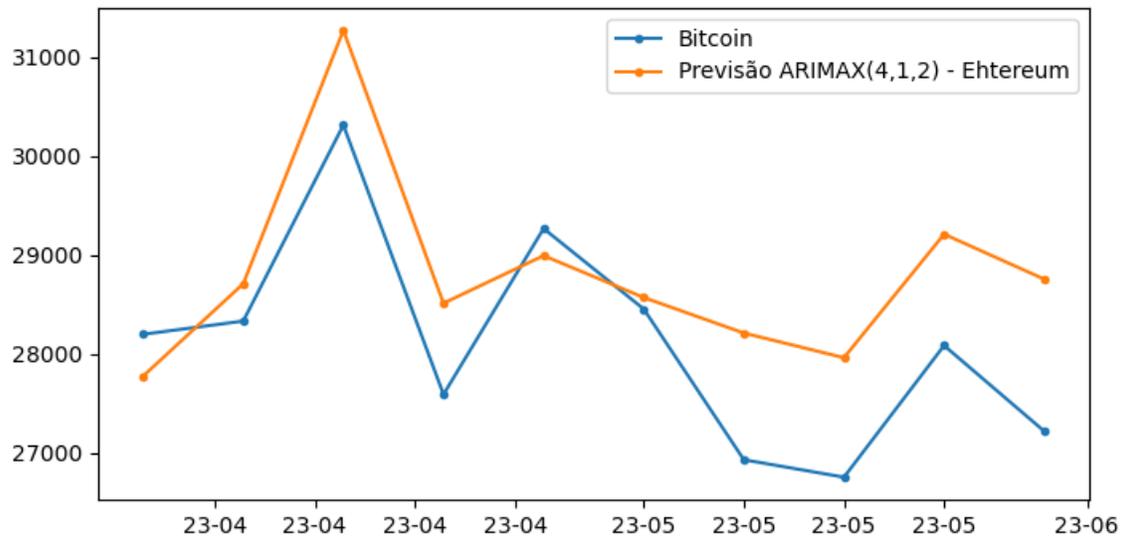


Figura C.4. Representação gráfica do conjunto de teste e valores preditos para o modelo ARIMAX()- Ethereum.

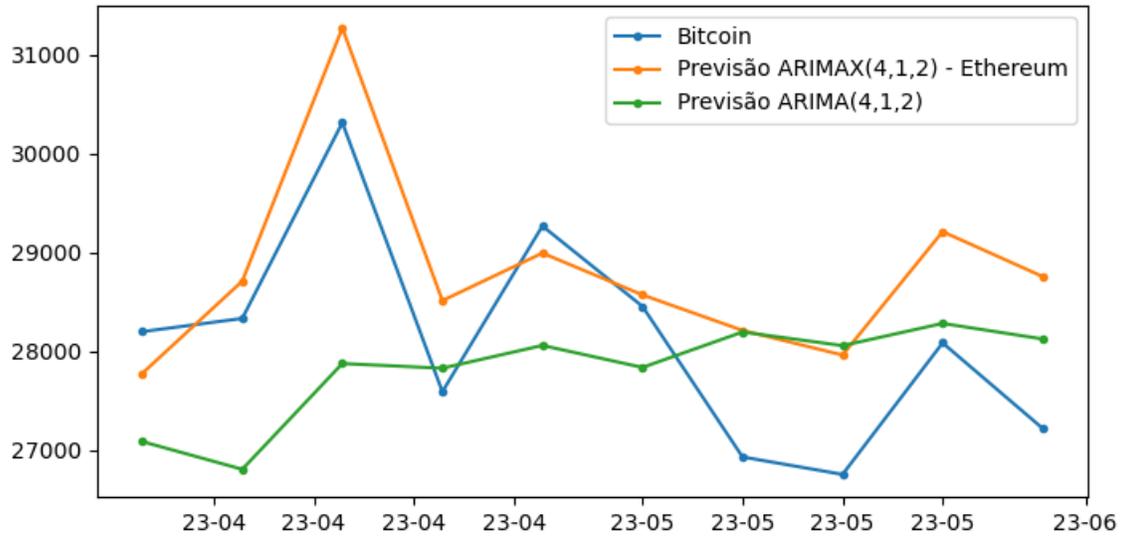


Figura C.5. Representação gráfica do conjunto de teste e valores preditos para o Modelos ARIMA 4,1,2) e ARIMAX()- Ethereum.

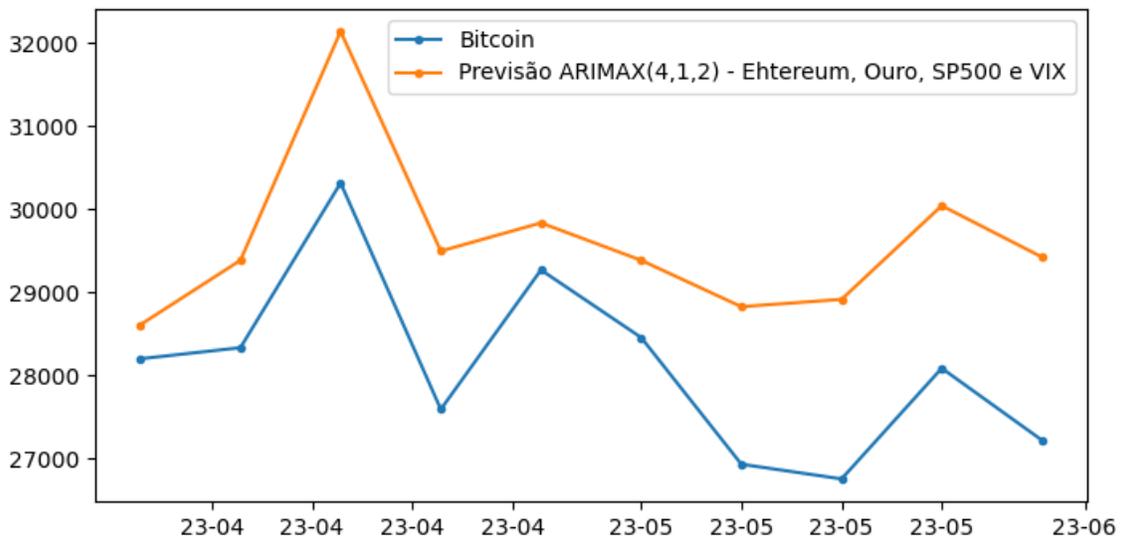


Figura C.5. Representação gráfica do conjunto de teste e valores preditos para o modelo ARIMAX()- Ethereum, Ouro, SP500 e VIX.