



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO
ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO
RELATÓRIO DE ESTÁGIO

**DESENVOLVIMENTO DE UM MODELO DE
PREDITIVO DO PREÇO DE ENERGIA NO MERCADO
IBÉRICO**

VITOR TRAVASSOS BARBOZA

OUTUBRO – 2022



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO

ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO

RELATÓRIO DE ESTÁGIO

DESENVOLVIMENTO DE UM MODELO DE PREDITIVO DO PREÇO DE ENERGIA NO MERCADO IBÉRICO

VITOR TRAVASSOS BARBOZA

ORIENTAÇÃO: NUNO SOBREIRA

OUTUBRO – 2022

Abstract

The objective of this work is to build accurate forecasts for the price of energy in the Iberian market. Good forecasts allow companies to make better decisions at the business level. To achieve this goal, I collected data which is relevant to develop forecasting models under an econometric framework.

The present work used the ARIMA and ARIMAX models to obtain forecasts. The Box-Jenkins methodology was used as a reference for the model building process. I used the R software to perform data analysis, time series modelling and compute the forecasts for the variable of interest.

I find that the ARIMAX models performed better than the ARIMA models in predicting the price of energy.

Keywords: Forecasting; Energy; ARIMA; ARIMAX; R; Box-Jenkins Methodology

Resumo

O presente trabalho tem como objetivo prever o preço de energia no mercado Ibérico. Um modelo de previsão assertivo permite que as empresas façam melhores decisões a nível empresarial. Para alcançar esse objetivo, foram obtidos os dados que serviram de base para o desenvolvimento dos modelos de previsão.

O presente trabalho utilizou os modelos ARIMA e ARIMAX para realizar as previsões. A metodologia Box-Jenkins foi utilizada como referência para a construção dos modelos. Todo o processo de análise e desenvolvimento dos modelos foi feito no software R.

Os modelos ARIMAX tiveram uma melhor performance em relação aos modelos ARIMA na previsão do preço de energia.

Palavras-chave: Previsão; Energia; ARIMA; ARIMAX; R; Metodologia Box-Jenkins

Agradecimentos

Primeiramente, gostaria de agradecer ao Professor Nuno Sobreira pela disponibilidade e por ter aceitado orientar este trabalho, pelo aconselhamento e pelas sugestões de melhoria. Gostaria de agradecer também a Professora Esmeralda Ramalho por todo o suporte durante esse período do mestrado.

Agradeço aos meus pais por todo o esforço e sacrifício que tiveram ao longo da vida para me proporcionarem a melhor educação possível. Agradeço a minha avó por todo cuidado e paciência que ela teve comigo ao longo dos anos.

Agradeço a todos os meus amigos que me ajudaram nessa etapa da minha vida, mas também queria agradecer os amigos presente em outras etapas da vida, pois sem vossa ajuda, não seria a pessoa que sou hoje.

INDÍCE

1.	Introdução	1
2	Revisão da literatura	4
2.1	Literatura genérica	4
2.2	O caso Português e o mecanismo Mibel	6
3	Dados e Metodologia	7
3.1	Dados	7
3.2	Metodologia e Estimação	9
3.2.1	Estacionaridade	10
3.2.2	Teste de Autocorrelação	12
3.2.3	ARIMA	13
3.2.4	ARIMAX	14
3.2.5	Erros de Previsão	15
3.2.6	Critérios de Seleção	16
4	Análise dos resultados	18
5	Conclusões	28
	Referências Bibliográficas	29
	Anexos – Figuras	32
	Anexos – Códigos em R	35

ÍNDICE DE TABELAS E FIGURAS

TABELAS

TABELA 1 – VALORES DE CORRELAÇÃO TESTE DE SPEARMAN	24
TABELA 2 – ERROS DE PREVISÃO DOS MODELOS	25

FIGURAS

FIGURA 1 – DIAGRAMA BOX-JENKINS	10
FIGURA 2 – GRÁFICO DO PREÇO MÉDIO SEMANAL DA ENERGIA EM PORTUGAL	18
FIGURA 3 – DECOMPOSIÇÃO SÉRIE PREÇO MÉDIO SEMANAL ENERGIA EM PORTUGAL.....	19
FIGURA 4 – GRÁFICOS DE FUNÇÃO DE AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL.....	20
FIGURA 5 – TESTE ADF COM A SÉRIE ORIGINAL	20
FIGURA 6 – TESTE ADF COM A SÉRIE DIFERENCIADA.....	21
FIGURA 7 – GRÁFICOS DE FUNÇÃO DE AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL COM A SÉRIE DIFERENCIADA ...	22
FIGURA 8 – PREVISÕES COM OS MODELOS ARIMAX	26
FIGURA 9 – COMPARAÇÃO DE VALORES COM AS PREVISÕES DO MODELO ARIMAX (2,1,3).....	27
FIGURA 10 – TESTE SHAPIRO WILK	32
FIGURA 11 – TESTE DE CORRELAÇÃO SPEARMAN.....	32
FIGURA 12– SÉRIE DE MODELOS ARIMA	32
FIGURA 13 – SÉRIE DE MODELOS ARIMAX.....	33
FIGURA 14 – TESTE LJUNG-BOX – MODELOS ARIMA.....	33
FIGURA 15– TESTE LJUNG-BOX – MODELOS ARIMAX.....	33
FIGURA 16– COEFICIENTES – MODELOS ARIMA	34
FIGURA 17– COEFICIENTES – MODELOS ARIMAX.....	34

1. Introdução

Até ao início dos anos 90, havia um monopólio no mercado de energia elétrica uma vez que a venda era realizada, quase integralmente, por empresas governamentais. De acordo com Weron (2014), a partir dessa altura houve o início do processo de desregulação em vários mercados de energia elétrica em todo o mundo, e a energia elétrica passou a ser negociada como uma *commodity*. Com isso, houve uma alteração no modelo de negócio no mercado elétrico com a introdução de uma maior oferta na aquisição da energia elétrica resultando em uma maior competitividade.

Uma das modalidades de comercialização da energia é através do Mercado Spot (Spot Market), modalidade que será o foco do presente trabalho. O mercado spot é um ambiente no qual todos os agentes podem negociar livremente a compra e a venda de energia elétrica através de contratos bilaterais firmados entre empresas ou operações. As operações no mercado spot corresponderão à comercialização dos montantes de energia não contratados bilateralmente ou eventuais necessidades de compra de energia por parte de algum agente para honrar seus contratos (Bettega & Ramos, 2001). Além disso, ao contrário de outras *commodities*, o mercado spot é tipicamente *day-ahead market*.

A eletricidade é uma *commodity* com características especiais. A primeira característica a se destacar é que a procura de eletricidade é inerente a questões climáticas (como temperatura e estação do ano). A segunda característica a se destacar é que a eletricidade não pode ser armazenada economicamente (Shahidehpour, 2002) e, para que exista estabilidade do sistema elétrico, é necessário um equilíbrio constante entre a produção e o consumo. A terceira característica é a existência de possíveis congestionamentos na rede de transmissão de energia, “os congestionamentos têm a sua

origem na ocorrência de restrições elétricas em linhas de transmissão que impedem a transferência mais económica de energia de um ponto ao outro do sistema respeitando-se o limite da capacidade máxima das linhas” (Silva, 2008).

Existem diversos mercados de energia no mundo, no entanto, o presente trabalho vai focar-se no Mercado Ibérico de Eletricidade (MIBEL). O MIBEL, resulta de um processo de cooperação desenvolvido pelos Governos de Portugal e Espanha com o objetivo de promover a integração dos sistemas eléctricos dos dois países. Os resultados que daí advieram constituíram um contributo significativo não só para a concretização do mercado de energia eléctrica a nível ibérico, mas também, à escala europeia, um passo significativo para a construção do Mercado Interno de Energia.

Dado o contexto na mudança no mercado de energia, assim como as características peculiares da energia eléctrica, existe por necessidade do desenvolvimento de um modelo preditivo de energia no Mercado Ibérico. Em outras palavras, o presente trabalho tem como objetivo prever o preço de energia a curto prazo, com a finalidade de fornecer informações que ajudarão na tomada de decisão a nível empresarial, visto que há uma grande competitividade neste mercado.

Por exemplo, com um modelo de previsão adequado para o preço de energia, as empresas terão acesso a mais informação para tomarem decisões mais adequadas na negociação de contratos de energia. Por exemplo, será melhor assinar um contrato de curto ou longo prazo com as empresas de energia? Este tipo de decisões pode ser relevante para a gestão financeira das empresas.

Esses questionamentos foram colocados em prática durante o estágio no qual participei na empresa Milestone Consulting. O período do estágio foi entre abril e agosto de 2021, e o estágio resultou neste presente trabalho.

A segunda secção do presente trabalho será reservada para a revisão na literatura sobre o tema do trabalho. A revisão da literatura é uma das partes estruturantes da dissertação e será onde o investigador perceberá qual caminho percorrer em busca do modelo de previsão do preço de energia.

A terceira secção apresentará a metodologia utilizada no trabalho e abordará os dados utilizados no trabalho.

A quarta secção será reservada para apresentação e análise de diferentes modelos para a previsão do preço de energia. Nesta secção, será feita a comparação dos diversos modelos criados ao longo do trabalho e serão utilizadas algumas métricas para definir a qualidade de previsão dos mesmos.

E por último, a secção 5 será a conclusão do trabalho, e para além das conclusões serão identificadas as limitações do estudo assim como eventuais recomendações de pesquisas complementares futuras.

2 Revisão da literatura

2.1 Literatura genérica

Existe um vasto número de trabalhos e artigos na literatura sobre a previsão do preço de energia. Estes trabalhos apresentam uma grande variedade de técnicas relevantes efeitos de previsão desta variável.

Primeiramente, é necessário entender a definição de *Electricity Price Forecasting* (EPF), ou previsão do preço de eletricidade. “A previsão dos preços da eletricidade (EPF) é um ramo da previsão de energia na interface entre econometria/estatística e engenharia, que se centra na previsão do local e preços a prazo nos mercados grossistas de eletricidade” (Weron & Ziel, 2008). Segundo os mesmos autores, a previsão do preço de energia começou na década de 1990 devido a uma desregulação que levou a aumento na competitividade nos mercados do Reino Unido e Escandinávia, e em seguida espalhou-se rapidamente pela Europa e América do Norte.

Que seja do meu conhecimento, um dos primeiros trabalhos com o objetivo de prever o preço de eletricidade foi desenvolvido por Wang & Ramsay (1997). A previsão foi realizada para o mercado de energia do Reino Unido e foi usada a técnica de Redes Neurais. O horizonte de previsão escolhido pelos autores foi o *day-ahead* (para o dia seguinte). O erro percentual médio absoluto (MAPE) foi utilizado para avaliar a qualidade das previsões e, de acordo com o MAPE, o modelo preditivo está bem ajustado aos dados.

Entre os vários trabalhos nesta área, destaca-se também o trabalho de Bunn (2000), que faz uma revisão dos principais aspectos metodológicos e de técnicas que se tornaram inovadoras para efeitos de previsão do volume de energia disponível na rede

assim como na previsão do preço da energia. A ênfase principal deste trabalho foi o cálculo de métodos intensivos, incluindo segmentação de variáveis, múltiplas modelações e combinação de técnicas de previsão.

Conejo et al. (2005) também estudaram algumas técnicas de previsão para o dia seguinte no mercado de energia elétrica. As técnicas consideradas incluíram análise de séries temporais e redes neurais. Em relação a abordagem de séries temporais, destaca-se o uso dos modelos ARIMA (autoregressivo, integrado e de média móvel) e de modelos dinâmicos para séries temporais. Os autores concluíram que as técnicas de séries temporais são mais eficazes do que as técnicas de redes neurais. Entre os modelos ARIMA e os modelos de regressão linear dinâmicos, estes últimos demonstraram melhor capacidade de previsão do que os modelos ARIMA.

Garcia et al. (2005) propuseram fazer a previsão do preço de energia utilizando os modelos GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*), metodologia que nesta altura já estava bastante desenvolvida na literatura sobre séries temporais. Os autores usaram este modelo para fazer a previsão no mercado de energia da Califórnia e da Espanha. Os autores concluíram que os modelos GARCH apresentavam tiveram uma melhor performance do que modelos ARIMA quando há bastante volatilidade nos dados.

Weron & Misiorek (2008) investigaram a qualidade de previsão de 12 métodos de séries temporais para previsão a curto prazo no mercado spot. Os métodos consideram o modelo padrão AR (autoregressivos) e suas extensões como, por exemplo, pré-processamento dos picos de preço, modelo autoregressivo semi-paramétrico, modelo de reversão para a média, e os modelos ARIMAX (que são os modelos ARIMA com a inclusão de variáveis exógenas). Os autores concluíram que os modelos que incluíram

variáveis exógenas tiveram uma qualidade de previsão melhor do que os outros métodos abordados no trabalho.

Amjady & Hemmati (2006) reviram alguns problemas presentes em EPF, mais especificamente, os problemas relacionados com as técnicas de séries temporais (AR, ARIMA, GARCH). Os autores argumentam que essas técnicas só têm sucesso em ocasiões nas quais a frequência dos dados é baixa, como padrões semanais. Os autores também argumentam que o uso de inteligência artificial e de abordagem híbrida (redes neurais, regressão fuzzy, entre outros) tem uma capacidade maior em lidar com comportamentos não lineares que podem estar presentes nos dados.

2.2 O caso Português e o mecanismo Mibel

Em relação a trabalhos relacionados com a previsão de energia no Mercado MIBEL, Dias (2009) utilizou os modelos ARIMA para realizar a previsão do preço de energia. O horizonte de previsão foi de hora em hora. O trabalho conseguiu captar o efeito semanal e produziu previsões com alguma precisão.

Ribeiro (2014) propôs calcular as previsões utilizando redes neurais artificiais. Outro contributo importante que o autor dá com o seu trabalho é o seguinte: “a volatilidade dos preços de eletricidade deve-se em grande parte à influência de variáveis meteorológicas, mais precisamente do vento”. O autor ainda diz que “cenários onde a produção eólica é elevada resultam num decréscimo dos preços de eletricidade e vice e versa”, e “a produção nuclear e hídrica tem grande influência na variabilidade dos preços MIBEL”.

3 Dados e Metodologia

3.1 Dados

Esta secção pretende fazer uma abordagem detalhada dos dados utilizados neste trabalho. Após a primeira etapa do estágio, que foi a realização de uma pesquisa na literatura em relação ao estado da arte da previsão do preço da eletricidade, construiu-se a base de dados que foi utilizada para o desenvolvimento dos modelos de previsão.

Os dados relativos ao Mercado Mibel foram obtidos no website do operador de mercado elétrico designado - OMIE. O OMIE é responsável pela gestão do mercado diário e intradiário de eletricidade na Península Ibérica (NEMO, de acordo com a terminologia europeia). Já os dados meteorológicos foram obtidos no website do *Weather Underground*, website que é um serviço comercial de meteorologia que fornece informações meteorológicas em tempo real pela Internet.

Os dados foram obtidos entre os dias 26 de abril de 2021 e 30 de abril de 2021. O período temporal utilizado vai entre os dias 01 de abril de 2018 e 31 de março de 2021, totalizando 1096 observações.

No site da OMIE, as variáveis obtidas foram:

- Preço médio de Portugal: preço médio diário em Portugal em Euros.
- Log. volume da energia total (MWh): volume diário de energia consumida.

Já no site do *Weather Underground*, as variáveis obtidas foram:

- Temperatura (° F)
- Velocidade do vento (mph)

Pretende-se utilizar estas variáveis meteorológicas como variáveis exógenas que serão acrescentadas nos modelos com o objetivo de testar se essas variáveis vão melhorar a capacidade preditiva dos modelos estatísticos desenvolvidos.

Além disso, a escolha das variáveis meteorológicas também tem como objetivo testar a relação do volume produzido de energia renováveis com o preço da energia. De acordo com Paraschiv et al. (2014), o aumento de consumo de energia a partir de fontes renováveis levou a uma diminuição parcial dos preços da eletricidade na Alemanha. As energias renováveis abordadas nesse trabalho são a energia solar e energia eólica. Note-se que a produção dessas energias está diretamente ligada à temperatura e à velocidade do vento, respectivamente.

Em relação à temperatura, coletou-se a temperatura diária de 3 cidades da Península Ibérica: Lisboa, Valencia e Alicante. Essas cidades foram escolhidas devido ao facto de usinas solares estarem localizadas nessas cidades ou nas proximidades. Após este processo de obtenção de dados e de modo a obter os dados finais que foram usados neste trabalho, calculou-se a temperatura média diária das cidades.

Em relação a velocidade do vento, obteve-se dados sobre a velocidade média diária de 7 cidades da Península Ibérica: Lisboa, Porto, Vigo, Salamanca, Badajoz, Zaragoza e Málaga. Tal justifica-se pelo facto de usinas eólicas estarem na proximidade dessas cidades. Após a obtenção desses dados, calculou-se a velocidade média diária do vento nessas cidades.

Por fim, todos esses dados foram tratados e compilados num ficheiro de Excel. A modelação econométrica dos dados bem como o cálculo das previsões foi realizada no *Software R*, em conjunto com o ambiente de programação *RStudio*.

3.2 Metodologia e Estimação

Após a pesquisa e revisão do estado da arte em relação a previsão do preço de energia, o presente trabalho decidiu construir os modelos de previsão utilizando a abordagem para a previsão de séries temporais. A análise e modelação será realizada utilizando o software R.

Utilizou-se o método de Box & Jenkins (1970) que é abordagem bastante utilizada para a construção de modelos para séries temporais univariadas que tem como objetivo encontrar o modelo com melhor ajustamento a série temporal. Esse método tornou-se bastante popular nos anos 1970 entre os acadêmicos (Makridakis e Hibon, 1997).

Segundo Box et al. (2015), as fases de construção de um modelo seguindo a metodologia Box & Jenkins são:

- Identificar o modelo que melhor se adapta a cada série. Para isso, é necessário analisar os seguintes pontos: estacionaridade, sazonalidade, verificar possíveis ordens do modelo a partir da função de autocorrelação e função de autocorrelação parcial. Nessa fase, um princípio muito importante é a parcimônia, que consiste na busca de um modelo com melhor ajuste e menor número possível de parâmetros;
- Estimação dos parâmetros do modelo: podemos utilizar método dos momentos, estimadores dos mínimos quadrados e estimadores máxima verossimilhança;
- A avaliação do diagnóstico do modelo. Essa fase é importante pelos seguintes motivos: avaliar a significância dos parâmetros do modelo, a estabilidade do modelo (se o modelo apresenta boas ou más previsões) e a

qualidade de ajustamento do modelo. Caso o modelo não apresente nenhum problema no diagnóstico, ele está pronto para ser usado. Caso contrário, se o modelo apresentar algum problema no diagnóstico, é necessário repetir o ciclo abaixo até a obtenção de um modelo que não apresente problemas no diagnóstico.



FIGURA 1 – Diagrama Box-Jenkins

3.2.1 Estacionaridade

Ao trabalhar com séries temporais, é importante assegurar que os dados sejam estacionários antes de proceder a qualquer tipo de análise (Granger & Newbold, 1974; Gujarati, 2006).

De acordo com Box et al. (2015), um processo é estritamente estacionário se as suas propriedades não forem afetadas por uma mudança de origem temporal, ou seja, se a

distribuição de probabilidade conjunta associada às observações m , $Z_{t_1}, Z_{t_2}, \dots, Z_{t_m}$, e para todo conjunto de momentos T_1, T_2, \dots, T_m é igual a distribuição conjunta m $Z_{t_1+k}, Z_{t_2+k}, \dots, Z_{t_m+k}$ para o conjunto de momentos $T_{1+k}, T_{2+k}, \dots, T_{m+k}$. Isto é, estacionariedade estrita implica que a distribuição de probabilidade conjunta seja constante ao longo do tempo, ou seja, que não dependa do instante t .

Ainda de acordo com Box et al. (2015), processo fracamente estacionário ou estacionário de segunda ordem é menos restritivo que o processo estritamente estacionário. Por exemplo, se a média e a matriz de variância-covariância forem constantes ao longo do tempo, é suficiente para garantir que o processo seja estacionário de segunda ordem:

- $E [Z_t] = \mu$
- $E [Z_t^2] = \sigma^2 < \infty$
- $\text{cov} [Z_t, Z_{t+k}] = \gamma_k$

Para transformar uma série não estacionária em estacionária, Hyndman & Athanasopoulos (2019) sugerem que se aplique a diferenciação para ajudar a estabilizar a média da série temporal, e também remover (ou reduzir) a tendência e a sazonalidade.

É possível identificar a estacionariedade de uma série temporal a partir dos gráficos de função de autocorrelação e função de autocorrelação parcial. Porém, existem uma variedade de testes para identificar se a série é estacionária ou não. Um dos testes para verificar a estacionariedade é Dickey-Fuller (DF). A abordagem do teste DF assume que os erros ε_t são não correlacionados, e a grande cerne do teste DF é verificar a existência de raiz unitária, se a série tiver raiz unitária, ela não é estacionária.

Ainda, é possível incluir um maior número de defasagem nas diferenças, e esse teste é conhecido como Dickey-Fuller Aumentado (ADF), e o teste ADF é baseado na seguinte regressão auxiliar:

$$\Delta Y_t = B_0 + B_1 t + \pi Y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta Y_{t-i} + \varepsilon_t \quad (1)$$

E as hipóteses são:

- $H_0: \pi = 0$, presença de raiz unitária, logo a série é não estacionária.
- $H_1: \pi < 0$, não tem raiz unitária, logo a série é estacionária.

3.2.2 Teste de Autocorrelação

Um dos pontos para avaliar o diagnóstico do modelo é verificar a existência de autocorrelação nos resíduos do modelo estimado. De acordo com Box & Pierce (1970), se o modelo for estimado apropriadamente, não existirá autocorrelação nos resíduos. Quando os resíduos do modelo não estão correlacionados entre si, significa que toda a informação presente na série temporal foi aproveitada pelo modelo. Caso verifique que os resíduos estão correlacionados entre si, significa que o modelo está mal especificado e que existem melhorias que possam ser feitas a fim de captar toda a informação presente na série temporal. Essas melhorias consiste em aumentar a ordem do modelo ou incluir variáveis *dummies* sazonais.

O teste que será usado no presente trabalho com o objetivo de testar a presença de autocorrelação nos resíduos é o teste de Ljung-Box. Este tem como hipótese nula a não existência de autocorrelação nos resíduos.

$$Q = T(T + 2) \sum_{k=1}^s r_k^2 / (T - K) \quad (2)$$

Se o valor de calculado de Q exceder o valor critico χ^2 com s graus de liberdade, pelo menos um dos valores r_k^2 (autocorrelação) é estatisticamente diferente de zero no nível de significância especificado, e há evidências para a rejeição da hipótese nula. S é o número de defasamentos a serem testados

3.2.3 ARIMA

O modelo Autoregressivo Integrado de Médias Móveis (ARIMA) é uma generalização dos processos Autorregressivos e de Médias Móveis (ARMA) que incorpora séries não estacionárias, e possui apenas uma variável e descreve o comportamento da série em termos dos seus valores passados. (Brockwell e Richard A. Davis, 2002).

Os modelos ARIMA são aplicados em casos nos quais há evidências de não estacionaridade, e ao aplicar a diferenciação, que corresponde a parte “integrada” do modelo, a série se torna estacionária.

Esses modelos são representados por ARIMA (p,d,q), onde p corresponde a ordem da parte autorregressiva, d corresponde ao número de diferenciações para tornar a série estacionária, e q corresponde a ordem da parte de médias móveis do modelo. Dito isso, os modelos ARIMA podem ser expressados pela seguinte equação:

$$\varphi(B)z_t = \phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t \quad (3)$$

Onde:

$$\phi(B) = 1 - \phi_1 - \phi_2 B^2 - \dots - \phi_p B^p \quad (4)$$

$$\theta(B) = 1 - \theta_1 - \theta_2 B^2 - \dots - \theta_q B^q \quad (5)$$

E segue as seguintes condições:

- $\phi(B)$ é chamado de operador autoregressivo, e é assumida a estacionaridade, isto é, as raízes $\phi(B) = 0$ estão fora do círculo unitário.
- $\theta(B)$ é chamado de operador de média móveis, e é assumida a invertibilidade, isto é, as raízes $\theta(B) = 0$ estão fora do círculo unitário.

Os métodos mais utilizados para estimar os parâmetros dos Modelos ARIMA são: método dos mínimos quadrados e método da máxima verossimilhança.

3.2.4 ARIMAX

Como alternativa aos modelo ARIMA, o presente trabalho abordará os modelos ARIMAX. Estes combinam os modelos ARIMA com a inclusão de variáveis exógenas (Benneth et al., 2014). O modelo ARIMAX pode ser expressado pela seguinte equação:

$$y_t = \rho + \sum_{i=1}^p \beta_i y_{t-i} + \sum_{k=1}^r \omega_k w_k + \sum_{j=1}^q (\theta_j e_{t-j}) + e_t \quad (6)$$

Onde: y_t é a variável dependente, β_i são os coeficientes da parte autoregressiva, ω_j representa as variáveis exógenas, r é o número de variáveis exógenas, θ_j representa a parte de médias móveis, e e_t representa os ruídos branco.

3.2.5 Erros de Previsão

Após a estimação dos modelos descritos acima, o próximo passo é verificar a qualidade de previsão desses modelos. E a acurácia de previsão pode ser determinada considerando novos dados que não foram utilizados na estimação do modelo (Hyndman & Athanasopoulos, 2019).

Para Enders (2014), como os valores futuros da série temporal é desconhecido, é necessário separar os dados disponíveis em duas porções: dados de treino e dados de teste. Os dados treinos serão utilizados na estimação dos modelos, enquanto que os dados de teste serão utilizados para avaliar a acurácia de previsão dos modelos.

O erro de previsão é a diferença entre o valor observado e o valor previsto. E pode ser escrito da seguinte forma:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T} \quad (7)$$

Onde y_{T+h} são os valores observados e $\hat{y}_{T+h|T}$ são os valores previstos

Existem vários tipos de erro de previsão, no entanto, o presente trabalho focará nos seguintes erros: Erro Quadrático Médio (MSE) e Erro Percentual Absoluto Médio (MAPE).

O MSE é um tipo de erro que depende da escala, logo, não pode ser usado para comparar séries que possuam unidades diferentes. E adicionalmente, o MSE é bastante sensível a erros elevados. O MSE é caracterizado pela seguinte expressão:

$$MSE = \frac{1}{H} \sum_{i=1}^H e_{1i}^2$$

(8)

Onde H é igual ao tamanho da amostra.

De acordo com Khair et al (2017), o MAPE é calculado utilizando o erro absoluto em cada período dividido pelos valores observados, depois é calculada a média dessas percentagens. Isto é, mede o tamanho do erro em termos percentuais, e é uma medida que não depende da escala dos dados. O MAPE é caracterizado pela seguinte expressão:

$$MAPE = \frac{1}{H} \sum_{t=1}^H \left| \frac{Y_T - \hat{Y}_T}{Y_T} \right|$$

(9)

3.2.6 Critérios de Seleção

Ao estimar um modelo, é necessário verificar a qualidade do ajuste aos dados. Ao adicionar novos parâmetros p e/ou q, haverá a redução da soma quadrada dos resíduos estimados. No entanto, também é necessário levar em consideração o princípio da parcimônia ao acrescentar novos parâmetros (Enders, 2014). Existem algumas métricas que mensuram a qualidade de um modelo estatístico visando a simplicidade, a que será abordada no presente trabalho é: Critério de Informação de Akaike (AIC) .

Essa métricas têm como o principal objetivo atingir um *trade-off* ótimo entre a redução da soma quadrada dos resíduos e a parcimônia do modelo. A métricas têm a seguinte fórmula:

$$AIC = T \ln(\text{sum of squared residuals}) + 2n$$

(10)

Onde n = número de parâmetros estimados; T = número de observações.

Ao selecionar o modelo de acordo com a métrica, sugere-se que a escolha seja feita de acordo com os menores valores de AIC.

4 Análise dos resultados

Devido ao grande volume de dados diários e a dificuldade de trabalhar com eles, resolveu-se fazer uma agregação semanal e, logo, houve uma redução de 1096 para 160 observações. Em relação ao horizonte de previsão, os modelos serão treinados para prever as próximas 8 semanas. Chegou-se aos dados semanais a partir da média dos valores diários por semana.

O primeiro passo será a realização de uma análise descritiva da série. Na figura abaixo, temos a série original do preço de energia.

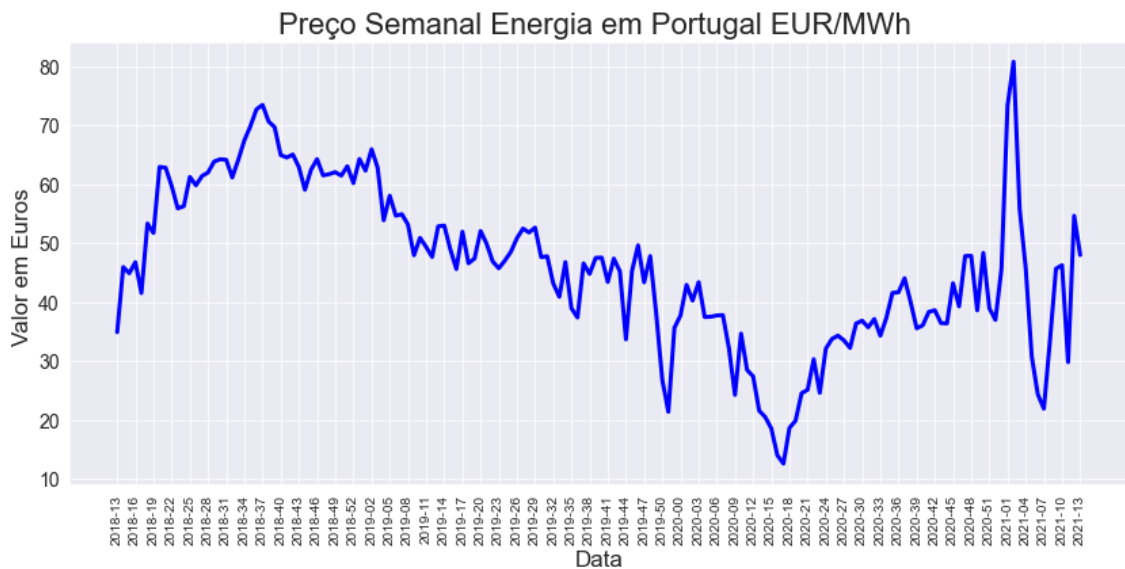


FIGURA 2 – Gráfico do Preço Médio Semanal da Energia em Portugal

Analisando a figura acima, há uma tendência crescente inicialmente, e após um tempo há uma reversão na tendência, tornando-se decrescente. Para um melhor entendimento do comportamento da série, realizou-se a decomposição da mesma. A decomposição de séries temporais é um processo estatístico que tem como objetivo separar uma série temporal em componentes distintos, tais como tendência, sazonalidade, ciclo e ruído aleatório. Esses componentes ajudam a compreender e

explicar a dinâmica da série temporal, além de permitir previsões futuras mais precisas. Utilizou-se uma função no R que realiza essa decomposição.

De acordo com a figura abaixo, percebe-se que a componente sazonal na série se repete a cada 52 semanas (aproximadamente 1 ano), e a componente sazonal se acentua, geralmente, no começo do ano, o que coincide com o inverno. E em relação a componente de tendência, podemos observar que a tendência apresenta um declive negativo, porém, há uma reversão na tendência a partir da metade do ano de 2020.

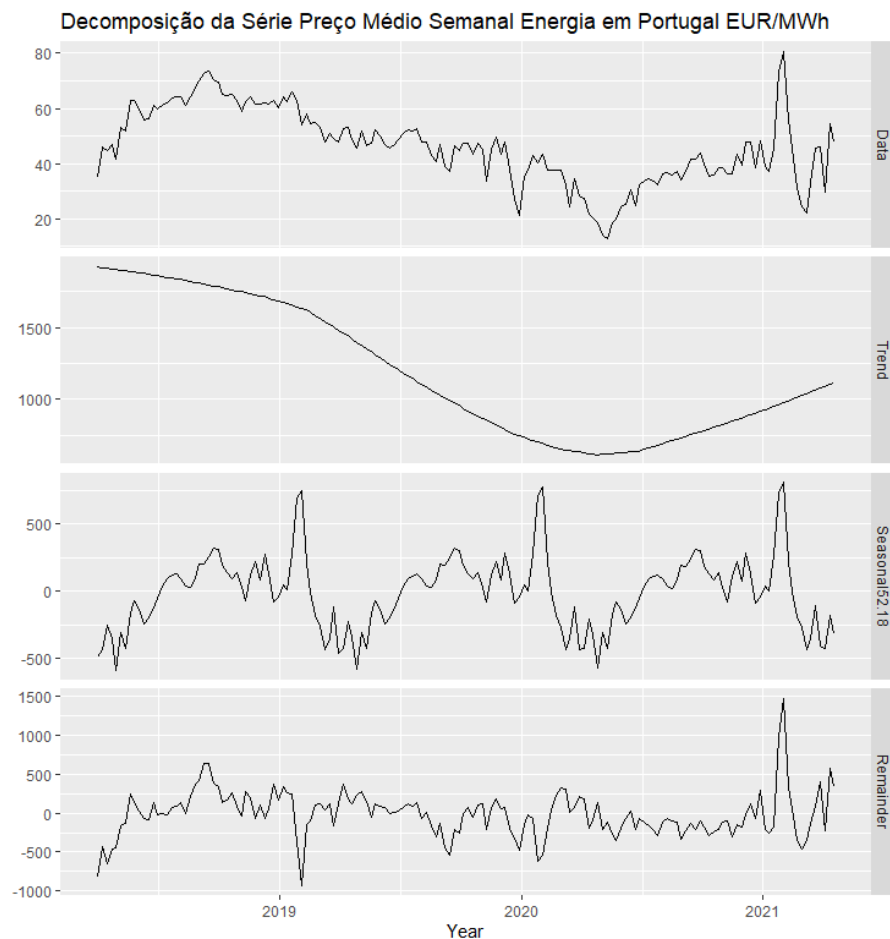


FIGURA 3 – Decomposição Série Preço Médio Semanal Energia em Portugal

O próximo passo será analisar os gráficos de função de autocorrelação e função de autocorrelação parcial, mas primeiro será feita uma transformação logarítmica para

diminuir os efeitos da variância. Os modelos utilizarão a série com transformação logarítmica.

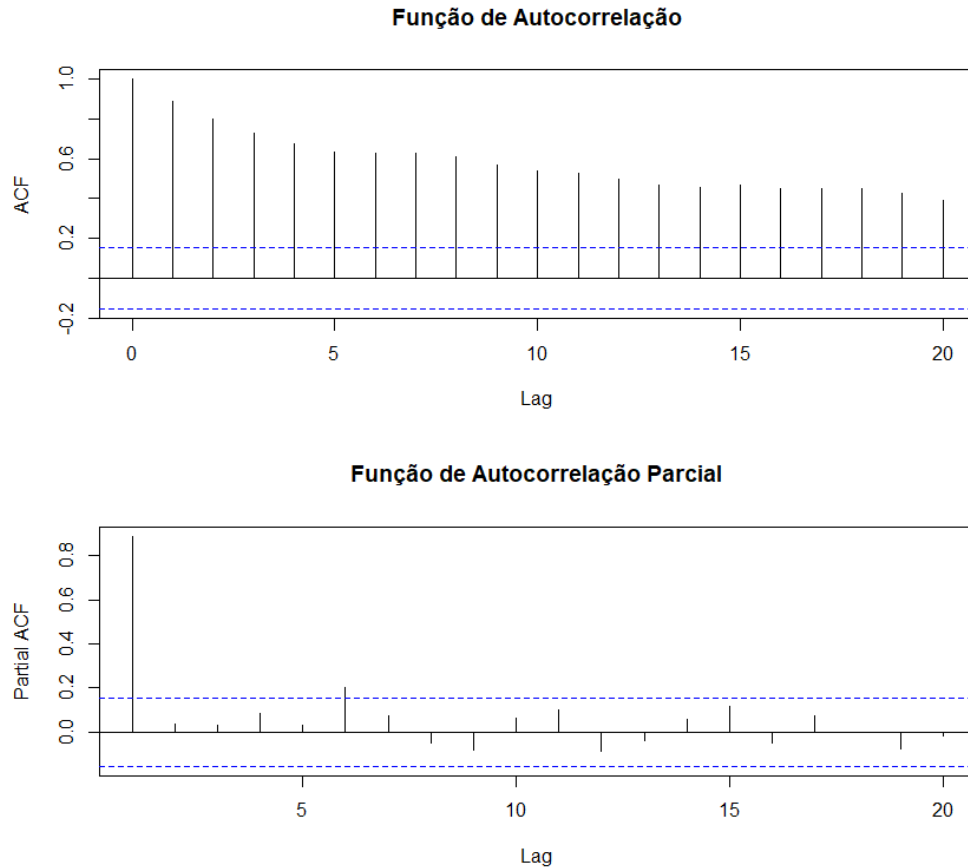


FIGURA 4 – Gráficos de Função de Autocorrelação e Autocorrelação Parcial

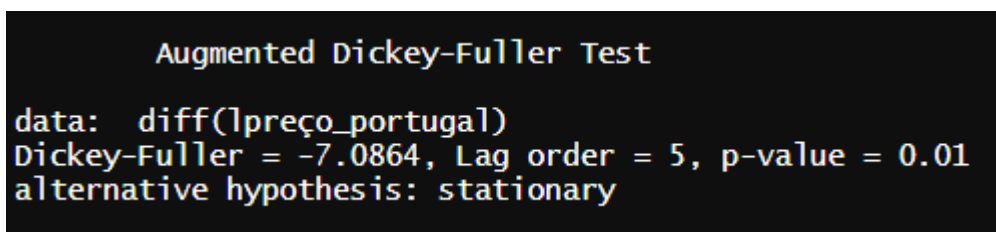
De acordo com o gráfico Função de Autocorrelação, percebe-se que o decair é muito lento, o que sugere que a série é não estacionária na média. Para confirmar se a série é estacionária ou não, realizou-se o teste ADF.

```
Augmented Dickey-Fuller Test
data: lpreço_portugal
Dickey-Fuller = -2.267, Lag order = 5, p-value = 0.465
alternative hypothesis: stationary
```

FIGURA 5 – Teste ADF com a Série Original

Com as hipóteses definidas pelo teste H_0 : série não é estacionária e H_1 : a série é estacionária. Logo sendo $p\text{-value}=0,46 > 0,05$; portanto não é possível rejeitar H_0 ao nível de 5% de significância e, logo, a série é não estacionária.

Com o objetivo de tornar a série estacionária, aplicou-se a diferenciação na série e em seguida realizou-se novamente o teste ADF.



```
Augmented Dickey-Fuller Test
data: diff(lpreço_portugal)
Dickey-Fuller = -7.0864, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

FIGURA 6 – Teste ADF com a Série Diferenciada

Após aplicar a diferenciação na série, houve a rejeição de H_0 , pois $p\text{-value}=0,01 < 0,05$; logo, a série é estacionária.

Posteriormente ao aplicar a diferenciação da série e ao averiguar que a série é estacionária, verificou-se novamente os gráficos de função de autocorrelação e função de autocorrelação parcial.

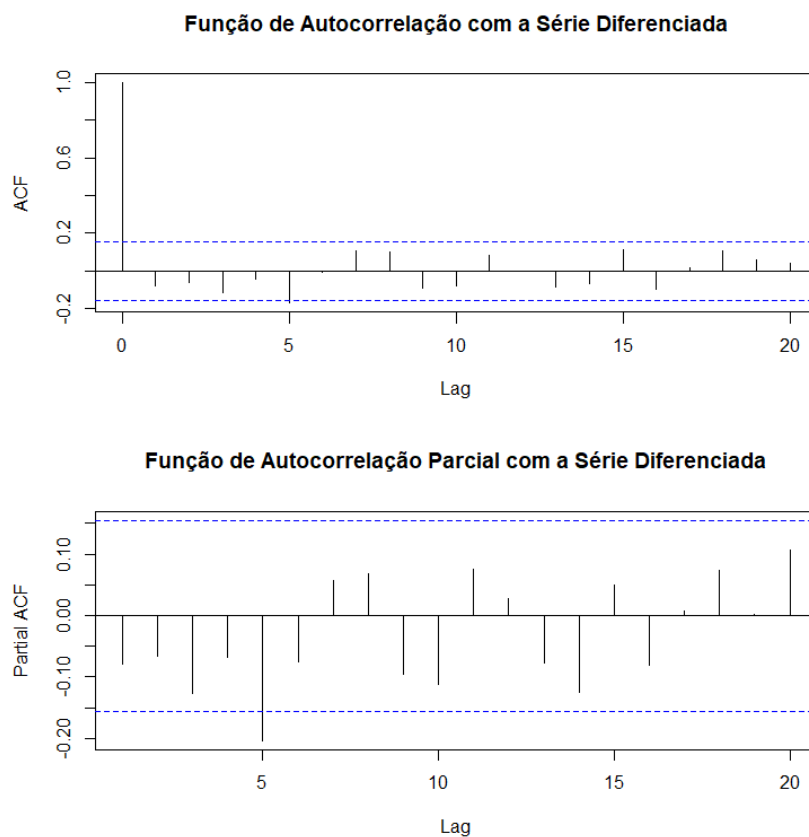


FIGURA 7 – Gráficos de Função de Autocorrelação e Autocorrelação Parcial com a Série Diferenciada

De acordo com a figura acima, percebe-se que a grande maioria dos valores de correlação estão próximas de zero e dentro do intervalo de confiança, logo, conclui-se que esses valores não são significativos.

Posteriormente, estimou-se uma série de modelos de diferentes ordens utilizando algumas funções do R, e houve a seleção de 4 modelos para realizar a comparação da qualidade de previsão dos respectivos modelos através das métricas de avaliação. Os modelos foram selecionados de acordo com o AIC.

Foram selecionados 2 modelos ARIMA e 2 modelos ARIMAX:

- ARIMA (2,1,2)
- ARIMA (1,1,0)
- ARIMAX (2,1,3)
- ARIMAX (2,1,2)

A ordem do modelo ARIMA (2,1,2) foi definida através de uma função presente em um *package* do R chamada *auto.arima()*, essa função retorna o modelo com o menor AIC. O modelo ARIMAX (2,1,3) também foi escolhido pelos mesmos critérios.

Para definir a ordem do modelo ARIMA(1,1,0), correu-se uma série de modelos e selecionou-se o modelo de acordo com menor AIC. Para definir a ordem do modelo ARIMAX(2,1,2), correu-se uma série de modelos e selecionou-se o modelo de acordo com menor AIC.

Em relação aos modelos ARIMAX, as seguintes variáveis exógenas estavam disponíveis para incluir no modelo: Log. Volume da energia total, temperatura média e velocidade média do vento. A justificativa da inclusão dessas variáveis é testar a hipótese de que as mesmas contribuem positivamente nas previsões. A escolha do Log. Volume da energia total é testar a hipótese se volume de energia consumido influencia diretamente no preço da energia. E a escolha das variáveis temperatura média e velocidade média do vento é testar a hipótese de que o preço da energia é afetado pelas questões meteorológicas. No entanto, para testar essas hipóteses e definir quais variáveis exógenas de fato serão incluídas, aplicou-se um teste de correlação e só serão incluídas as variáveis que apresentarem uma relação estatisticamente significativa com a variável dependente.

Primeiramente, verificou-se se essas variáveis seguem uma distribuição normal para definir qual teste de correlação utilizar. Utilizou-se o teste Shapiro-Wilk e a hipótese nula do teste é que a variável segue uma distribuição normal. Os outputs do teste podem ser visto para as respectivas variáveis no anexo. E houve rejeição da hipótese nula para todas as variáveis a nível de 5%, exceto a variável Log. Volume da energia total que segue uma distribuição normal, logo, será utilizado o teste de correlação de Spearman.

Variável	Valor de Correlação
Log. Volume da energia total	0,27**
Temperatura Média	-0,01
Velocidade média do Vento	-0,24**
** Estatisticamente significativo a 5%	

TABELA 1 – Valores de Correlação teste de Spearman

Na tabela acima estão os valores de correlação da variável preço médio de energia com as respectivas variáveis exógenas, e percebe-se que à medida que aumenta o volume de energia consumido, em média, há um aumento no preço de energia, sendo essa relação estatisticamente significativa ao nível de 5%. E à medida que a velocidade média do vento aumenta, em média, há diminuição no preço da energia, sendo essa relação estatisticamente significativa ao nível de 5%. E não há uma relação estatisticamente significativa entre preço de energia e temperatura média. Portanto, as variáveis Log. Volume da energia total e velocidade média do vento foram incluídas nos modelos ARIMAX.

Para aferir a qualidade de previsão dos modelos, é necessário comparar os valores de teste com os valores de previsão dos modelos. Para obter esses valores, houve a separação dos dados de treino e teste, 95% dos dados foram de treino e 5% dos dados foram de teste. Os modelos foram estimados com os dados de treino, e depois fez a

previsão para as próximas 8 semanas e comparou os valores previstos pelos modelos com os valores dos dados de teste.

No entanto, antes de calcular as previsões, é necessário analisar os resíduos dos modelos estimados, e será aplicado o teste Ljung-Box para detectar a presença de autocorrelação nos erros. Após isso, é que poderemos calcular os valores de previsão.

Aplicou-se o teste Ljung-Box nos resíduos para verificar se os modelos estão adequados corretamente. De acordo com o outputs dos respectivos modelos presente no anexo, percebe-se que não houve rejeição da hipótese nula ao nível de 5% de significância e, portanto, o modelos não apresentam problema nos diagnósticos.

Após verificar que os modelos estão adequados, é possível calcular as previsões dos valores futuros e escolher os melhores modelos de acordo com o MAPE e o MSE. Os valores para aferir a acurácia dos modelos está no quadro abaixo.

Modelo	MAPE	MSE
ARIMA (2,1,2)	7,41%	0,08
ARIMA (1,1,0)	8,45%	0,12
ARIMAX (2,1,3)	6,27%	0,07
ARIMAX (2,1,2)	6,98%	0,08

TABELA 2 – Erros de previsão dos modelos

De acordo com o quadro acima, percebe-se que os modelos nos quais tiveram os menores erros de previsão foram os modelos ARIMAX. E é importante destacar que as variáveis exógenas tiveram impacto positivos nas previsões. Os modelos ARIMAX serão escolhidos para as análises mais detalhadas, pois os mesmos apresentaram os menores erros de previsão. As previsões dos respectivos modelos estão ilustradas nos gráficos abaixo.

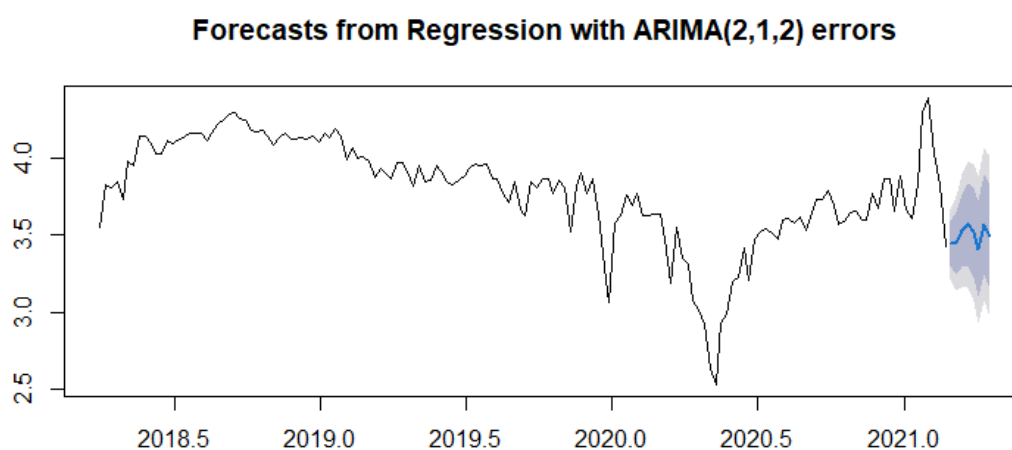
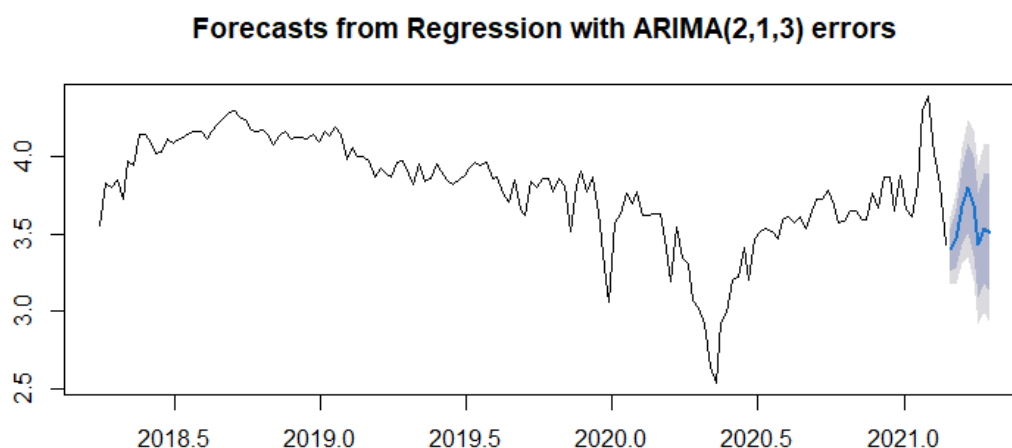


FIGURA 8 – Previsões com os Modelos ARIMAX

Ao analisar os gráficos de previsão dos respectivos modelos, nota-se que o modelo ARIMAX (2,1,3) sugere que haverá uma maior variação no preço da energia nas próximas 8 semanas do que o modelo ARIMAX (2,1,2). O cinza escuro nos gráficos representa o intervalo de confiança a 95% e o cinza claro representa o intervalo de confiança a 99%.

O modelo ARIMAX (2,1,3) será analisado mais detalhadamente, pois foi o modelo no qual teve os menores erros de previsão. Ao comparar os valores de previstos por este

modelo com os valores de teste, percebe-se as previsões estão coerentes, como pode ser visto no gráfico abaixo.

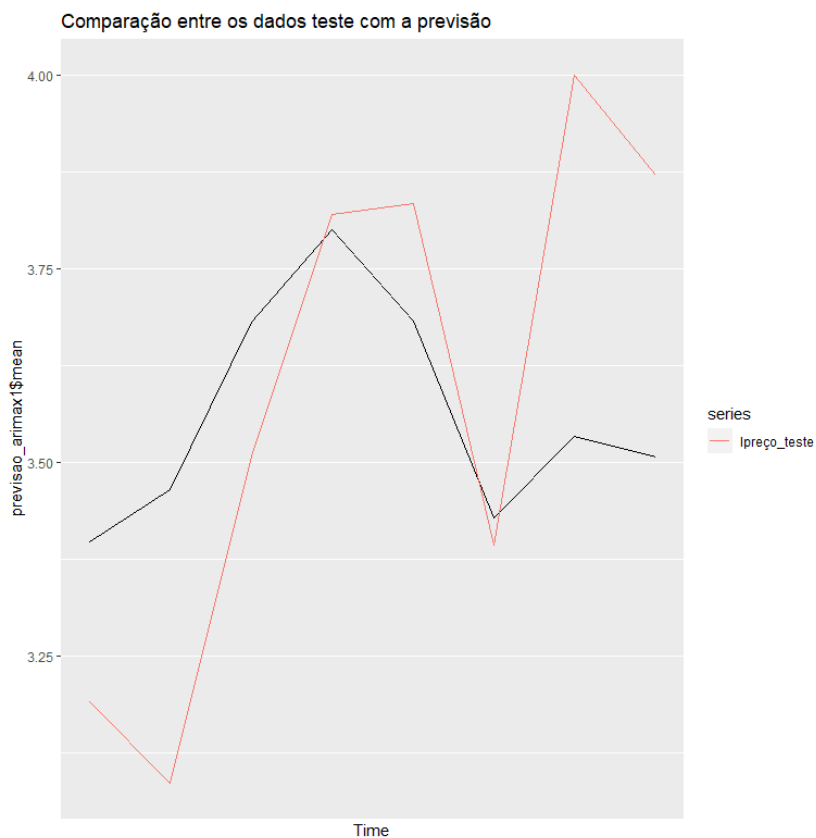


FIGURA 9 – Comparação de valores com as previsões do modelo ARIMAX (2,1,3)

5 Conclusões

Tendo em vista o crescimento do processo de desmonopolização no mercado de energia elétrica a partir dos anos 90 e a introdução de oferta maior de energia, é primordial que as empresas façam as suas decisões estratégicas baseadas em dados para se manterem competitivas.

Considerando do que foi exposto no presente trabalho, acredita-se que o objetivo principal que era estimar um modelo para prever o preço de energia a curto prazo, com a finalidade de fornecer informações que ajudarão na tomada de decisão a nível empresarial foi atingido.

Os modelos que tiveram os menores erros de previsão foram os modelos ARIMAX, o que converge com o estudo Weron & Misiorek (2008), no qual os autores chegaram à conclusão que a inclusão de variáveis exógenas melhorou qualidade de previsão dos modelos. Porém, os modelos ARIMAX apresentam algumas limitações, pois ao fazer a previsão para horizontes temporais, é necessário construir modelos adicionais para prever os valores futuros das variáveis exógenas que serão incluídas no modelo ARIMAX. Caso a previsão das variáveis exógenas não apresente uma qualidade de previsão satisfatória, as previsões dos modelos ARIMAX serão prejudicadas.

Para os próximos trabalhos, é interessante coletar uma amostra maior para tentar obter melhores resultados de previsão. Adicionalmente, uma possibilidade é em vez de trabalhar com dados semanais, é passar a trabalhar com dados diários. E por último, pode-se adotar uma nova abordagem para realizar as previsões que é utilizar as técnicas de redes neurais, como pode ser visto na literatura, apresentou melhores resultados em alguns casos.

Referências Bibliográficas

Amjady, N., & Hemmati, M. (2006). Energy price forecasting. IEEE Power and Energy Magazine, March/April, Pages 20–29.

Bennett, C.; Stewart, R.A.; Lu, J (2014). Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. Energies, v. 7, n. 5, Pages 2938-2960

Bettega, R., & Ramos, F. (2001). O Impacto da Comercialização no Mercado Spot de Energia na Análise de Viabilidade de Hidrelétricas Vol. 6, Nº 1, Pages 43-70

Brockwell, P. and R. Davis (1991). Time Series: Theory and Methods, 2nd Edition, New York: Springer-Verlag.

Box, G., & Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Box, G. P. E., Jenkins, G. M., Reinsel, G. C. e Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control, 5th Edition, Wiley.

Box, George E.P., and Pierce, David A. (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," Journal of the American Statistical Association, 65, Pages 1509-1526

Bunn, D. W. (2000). Forecasting loads and prices in competitive power markets. Proceedings of the IEEE, 88(2), Pages 163–169.

Conejo, A. J., Contreras, J., Espínola, R. & Plazas, M. A. (2005), "Forecasting electricity prices for a day-ahead pool-based electric energy market," Int. Jour. of Forecasting, vol. 21, no. 3, Pages 435–462, July-May 2005.

Dias, A. V. (2009), Forecasting Hourly Prices in the Portuguese Power Market with ARIMA Models. Dissertação de mestrado Instituto Universitário de Lisboa – ISCTE

Enders, W. (2014). Applied Econometrics Time Series, 4th Edition, Wiley.

Garcia, R., Contreras, J., van Akkeren, M., & Garcia, J. (2005), “A GARCH forecasting model to predict day-ahead electricity prices,” IEEE Transactions on Power Systems, vol. 20, no. 2, Pages 867–874, May 2005.

Granger, C.W J. & Newbold, P. 1974. ‘Spurious regressions in econometrics’. Journal of Econometrics, 2:111-120.

Greunen, J., Heymans, A., Heerden, C. & Vuuren G. (2014) The Prominence of Stationarity in Time Series Forecasting, Studies in Economics and Econometrics, 38:1, Pages 1-16,

Gujarati, D.N. 2006. ‘Essentials of econometrics’. New York: McGraw Hill.

Hyndman, R.J., & Athanasopoulos, G. (2019). Forecasting: principles and practice, 3rd Ed. Melbourne, Australia: OTexts [Online]. Available at: <https://otexts.com/fpp3/>

Khair, U. et al (2017) J. Phys.: Conf. Ser. 930 012002

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. (1992). ‘Testing for the Null of Stationarity Against the Alternative of a Unit Root’, Journal of Econometrics, 54, Pages 159-178.

Makridakis S. Hibon M. (1997). ARMA models and the Box–Jenkins methodology. Journal of Forecasting, 16(3), Pages 147–163.

Paraschiv, F., et al. (2014), The impact of renewable energies on EEX day-ahead electricity prices. Energy Policy, Volume 73, October 2014, Pages 196-210

Ribeiro, J. G. A. (2014), Previsão de preços de eletricidade para o mercado MIBEL. Dissertação de mestrado Faculdade Engenharia da Universidade do Porto.

Silva, J. O. F. Alocação de Custos de Congestionamento em Redes de Transmissão de Energia Elétrica Baseada no Princípio Usuário-Pagador. Brasília, 2008. Tese. (Doutorado em Engenharia Elétrica). Publicação PPGENE.TD – 028/2008. Departamento de Engenharia Elétrica, Faculdade de Tecnologia, Universidade de Brasília, 2008. Brasília - DF, 108p.

Wang, A. & Ramsay, B. (1997), “Prediction of system marginal price in the UK Power Pool using neural networks,” in Proc. Intl. Conf. on Neural Networks, vol. 4, 9-12 June 1997, Pages 2116–2120.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. Volume 30, Issue 4, October–December 2014, Pages 1030-1081

Weron, R., & Misiorek, A. (2008). Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. International Journal of Forecasting, 24, Pages 744–763

Weron, R. & Ziel, F. (2018). Electricity price forecasting. HSC Research Reports HSC/18/08, Hugo Steinhaus Center, Wroclaw University of Technology.

Anexos – Figuras

FIGURA 10 – Teste Shapiro Wilk

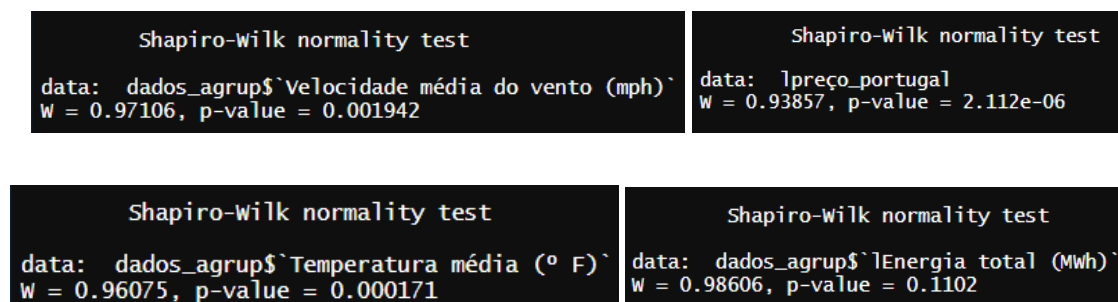


FIGURA 11 – Teste de Correlação Spearman

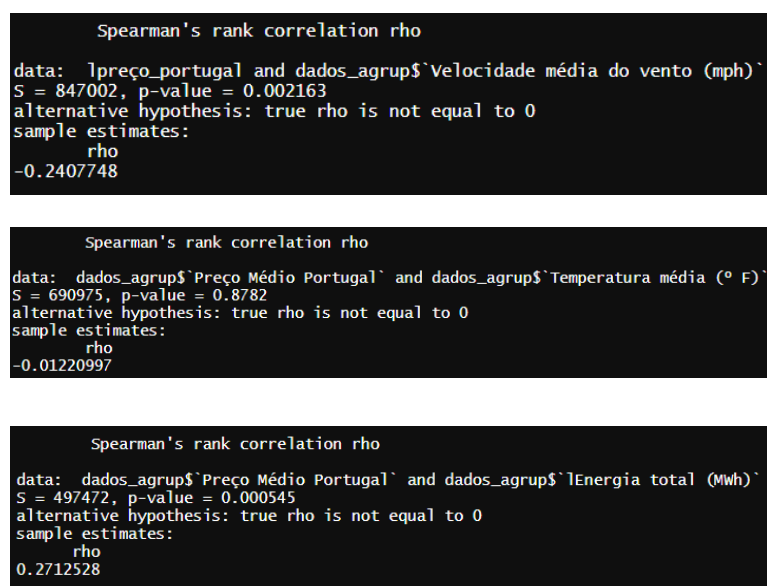


FIGURA 12 – Série de modelos ARIMA

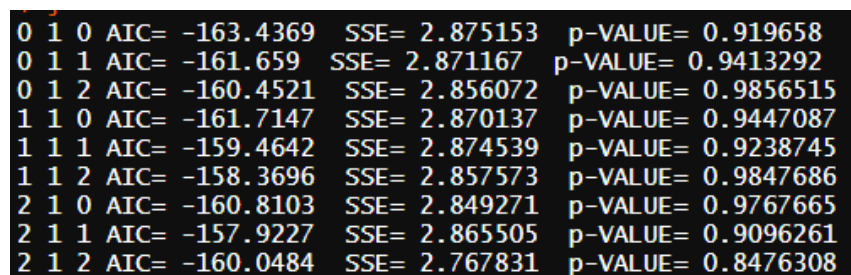


FIGURA 13 – Série de modelos ARIMAX

```

0 1 0 AIC= -220.0012  SSE= 1.927468  p-VALUE= 0.3267837
0 1 1 AIC= -218.0819  SSE= 1.926407  p-VALUE= 0.3425996
0 1 2 AIC= -220.097   SSE= 1.873864  p-VALUE= 0.7561616
0 1 3 AIC= -219.6353  SSE= 1.852907  p-VALUE= 0.7023634
1 1 0 AIC= -218.071   SSE= 1.92655   p-VALUE= 0.3406669
1 1 1 AIC= -218.855   SSE= 1.887297  p-VALUE= 0.6123448
1 1 2 AIC= -227.7495  SSE= 1.741615  p-VALUE= 0.5293614
1 1 3 AIC= -216.9036  SSE= 1.862494  p-VALUE= 0.6420301
2 1 0 AIC= -220.6063  SSE= 1.867389  p-VALUE= 0.7677981
2 1 1 AIC= -220.2907  SSE= 1.845271  p-VALUE= 0.7217103
2 1 2 AIC= -228.8313  SSE= 1.706921  p-VALUE= 0.9339457
2 1 3 AIC= -226.942   SSE= 1.705546  p-VALUE= 0.9679072
3 1 0 AIC= -219.0795  SSE= 1.861384  p-VALUE= 0.7709197
3 1 1 AIC= -216.6658  SSE= 1.866622  p-VALUE= 0.7410053
3 1 2 AIC= -223.5778  SSE= 1.743295  p-VALUE= 0.3459872
3 1 3 AIC= -225.4702  SSE= 1.698888  p-VALUE= 0.880506

```

FIGURA 14 – Teste Ljung-Box – Modelos ARIMA

Ljung-Box test	Ljung-Box test
data: Residuals from ARIMA(2,1,2)	data: Residuals from ARIMA(1,1,0)
Q* = 23.87, df = 26, p-value = 0.5834	Q* = 29.29, df = 29, p-value = 0.45
Model df: 4. Total lags used: 30	Model df: 1. Total lags used: 30

FIGURA 15 – Teste Ljung-Box – Modelos ARIMAX

```

Ljung-Box test
data: Residuals from Regression with ARIMA(2,1,3) errors
Q* = 30.049, df = 23, p-value = 0.148
Model df: 7. Total lags used: 30

```

```

Ljung-Box test
data: Residuals from Regression with ARIMA(2,1,2) errors
Q* = 25.232, df = 24, p-value = 0.3933
Model df: 6. Total lags used: 30

```


FIGURA 16 – Coeficientes – Modelos ARIMA

```

Series: lpreço_treino
ARIMA(2,1,2)

Coefficients:
      ar1      ar2      ma1      ma2
    -0.0035  0.6128  -0.1273  -0.6994
s.e.    0.6428  0.4748   0.6429   0.5554

sigma^2 estimated as 0.01955:  log likelihood=84.73
AIC=-159.46  AICc=-159.05  BIC=-144.38

```

```

Series: lpreço_treino
ARIMA(1,1,0)

Coefficients:
      ar1
    -0.0593
s.e.    0.0844

sigma^2 estimated as 0.01978:  log likelihood=82.44
AIC=-160.87  AICc=-160.79  BIC=-154.84

```

FIGURA 17 – Coeficientes – Modelos ARIMAX

```

Series: lpreço_treino
Regression with ARIMA(2,1,3) errors

Coefficients:
      ar1      ar2      ma1      ma2      ma3  Volume Energia  Velocidade Vento
    0.9576  -0.8774  -1.0278   1.1094  -0.2304           0.8551           -0.0458
s.e.    0.0863   0.0754   0.1330   0.0934   0.1206           0.3402           0.0045

sigma^2 estimated as 0.01209:  log likelihood=121.87
AIC=-227.75  AICc=-226.74  BIC=-203.61

```

```

Series: lpreço_treino
Regression with ARIMA(2,1,2) errors

Coefficients:
      ar1      ar2      ma1      ma2  Volume Energia  Velocidade Vento
    0.0746  0.7547  -0.1391  -0.8067           1.3074           -0.0453
s.e.      NaN      NaN      NaN      NaN           0.3833           0.0052

sigma^2 estimated as 0.01302:  log likelihood=116.43
AIC=-218.86  AICc=-218.07  BIC=-197.73

```

Anexos – Códigos em R

Bibliotecas

```
library(readxl)
library(astsa)
library(forecast)
library(fUnitRoots)
library(timeDate)
library(tseries)
library(xts)
library(lubridate)
library(ggplot2)
library(tidyverse)
library(dplyr)
library(MLmetrics)
library(writexl)
library(dummies)
library(strucchange)
```

Dados

```
dados_agrup <- read_xlsx('dados_agrup_semana.xlsx')
```

```
preco = ts(dados_agrup$`Preço Médio Portugal`,
           freq=365.25/7,
           start=decimal_date(ymd("2018-04-01")))
```

```
preço_treino <- head(preco, round(length(preco) * 0.95))
```

```
h <- length(preco) - length(preço_treino)
```

```
preço_teste <- tail(preco, h)
```

```
lpreço_portugal <- log(preco)
```

```
lpreço_treino <- head(lpreço_portugal, round(length(lpreço_portugal) * 0.95))
h <- length(lpreço_portugal) - length(lpreço_treino)
lpreço_teste <- tail(lpreço_portugal, h)
```

Exploratory Data Analysis

```
autoplot(preco) +
  ggtitle("Preço Semanal Energia em Portugal EUR/MWh ") +
  xlab("Year") +
  ylab("Valor em Euros")
```

```
mstl(preco, lambda = "auto") %>% autoplot()+
  ggtitle('Decomposição da Série Preço Médio Semanal Energia em Portugal
EUR/MWh')+
  xlab('Year')
```

Gráficos de Autocorrelação e Estacionaridade

```
par(mfrow=c(2:1))
acf(coredata(lpreço_portugal), main="Função de Autocorrelação", 20)
pacf(coredata(lpreço_portugal),main="Função de Autocorrelação Parcial", 20)
```

```
adf.test(lpreço_portugal)
```

```
acf(coredata(diff(lpreço_portugal)), main="Função de Autocorrelação com a Série
Diferenciada", 20)
pacf(coredata(diff(lpreço_portugal)),main="Função de Autocorrelação Parcial com a
Série Diferenciada", 20)
```

```
acf(diff(lpreço_portugal), main="Função de Autocorrelação com a Série Diferenciada",
20)
pacf(coredata(diff(lpreço_portugal)),main="Função de Autocorrelação Parcial com a
Série Diferenciada", 20)
```

```
adf.test(diff(lpreço_portugal))
```

```
##### Teste de normalidade de correlação #####
```

```
shapiro.test(lpreço_portugal)
```

```
shapiro.test(dados_agrup$`Velocidade média do vento (mph)`)
```

```
shapiro.test(dados_agrup$`Temperatura média (° F)`)
```

```
shapiro.test(dados_agrup$`IEnergia total (MWh)`)
```

```
cor.test(dados_agrup$`Preço Médio Portugal`, dados_agrup$`Velocidade média do  
vento (mph)`, method="spearman")
```

```
cor.test(lpreço_portugal, dados_agrup$`Velocidade média do vento (mph)`,  
method="spearman")
```

```
cor.test(dados_agrup$`Preço Médio Portugal`, dados_agrup$`Temperatura média (° F)`,  
method="spearman")
```

```
cor.test(dados_agrup$`Preço Médio Portugal`, dados_agrup$`IEnergia total (MWh)`,  
method="spearman")
```

```
##### Modeling #####
```

```
d=1
```

```
for(p in 1:3){
```

```
  for(q in 1:3){
```

```
    if(p+d+q<=11){
```

```
      model<-arima(x=lpreço_treino,order = c((p),d,(q)))
```

```
      pval<-Box.test(model$residuals, lag=log(length(model$residuals)))
```

```
      sse<-sum(model$residuals^2)
```

```
      cat(p-1,d,q-1, 'AIC=', model$aic, ' SSE=',sse,' p-VALUE=', pval$p.value,'\n')
```

```
    }
```

```
  }
```

```

## Model 1 ##
model_arima1 <- auto.arima(lpreço_treino) #(2,1,2)
model_arima1
checkresiduals(model_arima1)

for (h in 1:100){
  #h#
  show(h)
  show(Box.test(model_arima1$residuals, lag = h, type = "Ljung-Box", fitdf = 0))}

previsao_arima1 <- forecast(model_arima1, h=8)
MAPE(previsao_arima1$mean, lpreço_teste)*100
MSE(previsao_arima1$mean, lpreço_teste)

## Model 2 ##
model_arima2 <- Arima(lpreço_treino, order=c(1,1,0)) #(2,1,0)
model_arima2
checkresiduals(model_arima2)

for (h in 1:100){
  #h#
  show(h)
  show(Box.test(model_arima2$residuals, lag = h, type = "Ljung-Box", fitdf = 0))}

previsao_arima2 <- forecast(model_arima2, h=8)
previsao_arima2

MAPE(previsao_arima2$mean, lpreço_teste)*100
MSE(previsao_arima2$mean, lpreço_teste)

## Preparação Arimax ##

```

```
reg <- cbind(dados_agrup$`Energia total (MWh)`,dados_agrup$`Velocidade média do
vento (mph)`)
```

```
colnames(reg) <- c("Volume Energia", "Velocidade Vento")
```

```
reg_treino <- reg[1:152,]
```

```
reg_teste <- reg[153:160,]
```

```
for(p in 1:4){
```

```
  for(q in 1:4){
```

```
    if(p+d+q<=11){
```

```
      model<-arima(x=lpreço_treino, xreg= reg_treino,
```

```
                  order = c((p),1,(q)))
```

```
      pval<-Box.test(model$residuals, lag=log(length(model$residuals)))
```

```
      sse<-sum(model$residuals^2)
```

```
      cat(p-1,d,q-1, 'AIC=', model$aic, ' SSE=',sse, ' p-VALUE=', pval$p.value,'\n')
```

```
    }
```

```
  }
```

```
}
```

```
## Model 3 ##
```

```
model_arimax1 <- auto.arima(lpreço_treino,xreg=reg_treino,
```

```
                        stepwise=FALSE,approx=FALSE)
```

```
model_arimax1
```

```
checkresiduals(model_arimax1)
```

```
for (h in 1:100){
```

```
  #h#
```

```
  show(h)
```

```
  show(Box.test(model_arimax1$residuals, lag = h, type = "Ljung-Box", fitdf = 0))}
```

```
previsao_arimax1 <- forecast(model_arimax1, h=8,xreg=reg_teste)
```

```

MAPE(previsao_arimax1$mean, lpreço_teste)*100 #10.35%
MSE(previsao_arimax1$mean, lpreço_teste) # 0.44

## Model 4 ##
model_arimax2 <- Arima(lpreço_treino, order=c(2,1,2),xreg=reg_treino)
model_arimax2
checkresiduals(model_arimax2)

for (h in 1:100){
  #h#
  show(h)
  show(Box.test(model_arimax2$residuals, lag = h, type = "Ljung-Box", fitdf = 0))}

previsao_arimax2 <- forecast(model_arimax2, h=8,xreg=reg_teste)
plot(previsao_arimax1)
plot(previsao_arimax2)

MAPE(previsao_arimax2$mean, lpreço_teste)*100 #10.27%
MSE(previsao_arimax2$mean, lpreço_teste) # 0.34

plot(previsao_arimax1)
plot(previsao_arimax2)

autoplot(previsao_arimax1$mean)+autolayer(lpreço_teste)+
  ggtitle('Comparação entre os dados teste com a previsão')

autoplot(previsao_arimax1)+autolayer(lpreço_teste)+
  ggtitle('Comparação entre os dados teste com a previsão')

```