



UNIVERSIDADE DE ÉVORA

**ESCOLA DE CIÊNCIAS E TECNOLOGIA**

DEPARTAMENTO DE MATEMÁTICA

**Estudo dos Impactos da Política de Fixação do  
Preço dos Combustíveis em São Tomé e Príncipe**

Nome do Mestrando | Hearton Daunetty Amado de Menezes

Orientação | Professora Doutora Dulce Maria de Oliveira Gomes

**Mestrado em Matemática e Aplicações**

Área de Especialização | Matemática e Aplicações

Dissertação

Évora, 2019

**ESCOLA DE CIÊNCIAS E TECNOLOGIA**

**DEPARTAMENTO DE MATEMÁTICA**

**Estudo dos Impactos da Política de Fixação do  
Preço dos Combustíveis em São Tomé e Príncipe**

Nome do Mestrando | Hearton Daunetty Amado de Menezes

Orientação | Professora Doutora Dulce Maria de Oliveira Gomes

**Mestrado em Matemática e Aplicações**

Área de Especialização | Matemática e Aplicações

Dissertação

Évora, 2019

## **Dedicatória**

A Deus, porque sem ele nada seria possível;  
A Maria de Madre Deus Nazaré, minha saudosa avó. Tudo o que hoje sou, devo-o a esta grande  
mulher;  
Á minha família.



## **Agradecimentos**

Agradeço a todos que direta ou indiretamente participaram da elaboração deste trabalho.

Aos meus parentes mais próximos, minha avó, minha mãe, minha esposa e minhas filhas, meus tios pelo amor e paciência nas horas mais difíceis.

À Prof. Dra. Dulce Gomes pelo conhecimento, colaboração e empenho em tornar este trabalho uma realidade.

A Universidade de Évora pela oportunidade de realizar essa formação.

A empresa ENCO, SARL, pelo dados disponibilizados.

Aos colegas do mestrado, pela ajuda durante este período.

Peço desculpas se nesta hora esqueço de alguém, mas saiba que você também foi muito importante.



# Resumo

A flutuação do preço do petróleo e a sua subvenção praticada por alguns países, incluindo São Tomé e Príncipe, é um tema de grande relevância dada a enorme importância que o petróleo assume nas economias dos países que, tal como o nosso, não dispõem de recursos energéticos próprios (salvo a água, o vento e o sol).

A Empresa Nacional de Combustíveis e Óleo (ENCO, SARL), representa o monopólio de importação e venda dos produtos petrolíferos (gasolina, gasóleo, Jet-A1 e Petróleo) em São Tomé e Príncipe. A mesma é encarregada de elaborar a estrutura de preços dos produtos acima mencionados sempre que há uma importação. A elaboração desta estrutura de preços é baseada nas leis e normas de importação e vendas de produtos petrolíferos vigentes no País, onde o preço de venda ao consumidor final é fixo causando com isto um diferencial de preço que será o nosso objecto de estudo.

Pretendemos com esta dissertação identificar os beneficiários da política de fixação do preço dos produtos petrolíferos praticado pelo Estado santomense durante os anos de 2001 a 2017 e construir modelos estatísticos que permitam explicar a evolução desse diferencial e fazer uma previsão futura desse mesmo diferencial.

No final apresentaremos as conclusões e, caso se justifique, algumas recomendações futuras.

**Palavras-chave:** Produtos petrolíferos, estrutura de preços dos produtos petrolíferos, séries temporais.





# Abstract

## STUDY OF THE IMPACTS OF FUEL PRICE FIXING POLICY IN SÃO TOMÉ E PRÍNCIPE

The fluctuations in oil prices and its subsidy practiced by some countries, including São Tomé e Príncipe, is a subject of great importance given the enormous importance of oil in the economies of countries that, like ours, do not have their own energy resources (except water, wind and sun).

The National Fuel and Oil Company (ENCO) represents the monopoly on the import and sale of petroleum products (gasoline, diesel, Jet-A1 and Petroleum) in São Tomé e Príncipe, which is responsible for elaborating the price structure of the products mentioned above whenever there is an import. The elaboration is made based on the laws and norms of importation and sales of petroleum products in force in the Country, where the sale price to the final consumer is fixed causing with this a differential of price that will be our object of study.

We intend with this dissertation to identify the beneficiaries of the petroleum products pricing policy practiced by the State of São Tomé e Príncipe during the years 2001 to 2017, and to construct statistical models that allow to explain the evolution of this differential and to make a future prediction of this same differential.

In the final chapter, we will present the conclusions and, if appropriate, some future recommendations.

**Keywords:** Petroleum products, structure of prices of petroleum products, time series.



# Índice

<b>1</b>	<b>Introdução</b>	<b>17</b>
1.1	Motivação do estudo . . . . .	18
1.2	Breve história das Séries Temporais e alguns exemplos de aplicação . . . . .	19
1.2.1	Breve história das Séries Temporais . . . . .	19
1.2.2	Alguns exemplos de aplicação . . . . .	19
1.3	Organização da tese . . . . .	20
<b>2</b>	<b>Contextualização do caso de estudo</b>	<b>21</b>
2.1	Introdução . . . . .	22
2.2	Caracterização da ENCO, SARL . . . . .	22
2.3	Mapa de estrutura de preço . . . . .	23
2.4	Mercado dos combustíveis . . . . .	23
2.4.1	Mercado global . . . . .	23
2.4.2	Mercado nacional . . . . .	27
<b>3</b>	<b>Métodos quantitativos estocásticos de ajustamento e de previsão</b>	<b>30</b>
3.1	Introdução . . . . .	31
3.2	Séries temporais: breve introdução . . . . .	31
3.2.1	Conceitos gerais sobre processos estocásticos . . . . .	32
3.2.2	Funções de autocovariância, autocorrelação e autocorrelação parcial . . . . .	34
3.2.3	Estimação da média, variância e funções de autocovariância, autocorrelação e autocorrelação parcial . . . . .	35
3.3	Modelos para séries temporais lineares estacionárias . . . . .	36
3.3.1	Modelos autoregressivos . . . . .	36
3.3.2	Modelos de médias móveis . . . . .	38
3.3.3	Modelos autoregressivos e de médias móveis . . . . .	40
3.4	Modelos para séries temporais lineares não estacionárias . . . . .	41
3.4.1	Transformações para estabilizar a variância . . . . .	42
3.4.2	Transformações para eliminar as tendências . . . . .	43
3.4.3	Modelo ARIMA . . . . .	43
3.4.4	Modelo SARIMA . . . . .	44
3.5	Modelação de séries temporais . . . . .	45
3.5.1	Identificação do modelo . . . . .	45
3.5.2	Estimação dos parâmetros do modelo . . . . .	46
3.5.2.A	Estimador dos momentos para o modelo AR(p) . . . . .	46
3.5.2.B	Estimador dos mínimos quadrados condicionais para o modelo AR(p) . . . . .	47
3.5.2.C	Estimador de máxima verosimilhança condicionada para o modelo AR(p) . . . . .	47
3.5.3	Avaliação do diagnóstico . . . . .	48
3.5.3.A	Significância estatística dos parâmetros do modelo . . . . .	48
3.5.3.B	Estacionariedade e invertibilidade do modelo . . . . .	48
3.5.3.C	Estabilidade do modelo . . . . .	49
3.5.3.D	Redundância dos parâmetros do modelo . . . . .	49
3.5.3.E	Avaliação da qualidade do modelo . . . . .	49
3.5.4	CrITÉrios de selecção de modelos . . . . .	50
3.6	Previsão de séries temporais . . . . .	51
3.6.1	Performance preditiva . . . . .	52

<b>4</b>	<b>Aplicação ao caso de estudo</b>	<b>53</b>
4.1	Introdução . . . . .	54
4.2	Tratamento prévio dos dados . . . . .	54
4.3	Análise descritiva dos dados . . . . .	54
4.4	Modelação das séries . . . . .	58
4.4.1	Modelação dos diferenciais de preço da gasolina . . . . .	58
4.4.2	Modelação dos diferenciais de preço do gasóleo . . . . .	62
4.4.3	Modelação dos diferenciais de preço do petróleo . . . . .	66
4.5	Beneficiário da política de controlo do preço de combustível 2001-2017 . . . . .	71
<b>5</b>	<b>Conclusões</b>	<b>73</b>
<b>6</b>	<b>Anexos</b>	<b>76</b>
6.1	Anexos . . . . .	77
	<b>Referências bibliográficas</b>	<b>81</b>

# Lista de Figuras

2.1	Estrutura dos acionistas (Relatório e conta da ENCO, SARL 2017) . . . . .	22
2.2	Mapa de Estrutura de Preço (ENCO, SARL) . . . . .	24
2.3	Análise comparativa, entre a IEF e a OPEC, sobre a procura de petróleo (Relatório mensal de Abril de 2018) ( <a href="https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf">https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf</a> ) . . . . .	25
2.4	Análise comparativa, entre a IEF e a OPEC, sobre a oferta de petróleo (Relatório mensal de Abril de 2018) ( <a href="https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf">https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf</a> ) . . . . .	26
2.5	Mapa de estrutura de mercado (Hearton De Menezes, 2017) . . . . .	27
3.1	Série temporal com tendência crescente até 2008 (FRED government work <a href="https://fred.stlouisfed.org/series/GASREGCOVM">https://fred.stlouisfed.org/series/GASREGCOVM</a> ) . . . . .	32
3.2	Série temporal com sazonalidade (Samia, Keimara e Pedro, 2017 <a href="https://online.unisc.br/seer/index.php/cepe/article/download/9063/6870">https://online.unisc.br/seer/index.php/cepe/article/download/9063/6870</a> ) . . . . .	32
3.3	Série temporal, FAC e FACP de um ruído branco . . . . .	36
3.4	Série temporal, FAC e FACP de um AR(2) ( <a href="https://www.slideshare.net/SrinivasKNS1/presentation-42950288">https://www.slideshare.net/SrinivasKNS1/presentation-42950288</a> ) . . . . .	39
3.5	Série temporal, FAC e FACP de um MA(2) (dados simulados) . . . . .	40
3.6	Série temporal, FAC e FACP de um ARMA(1,1) ( <a href="https://www.xycoon.com/arma11_process.htm">https://www.xycoon.com/arma11_process.htm</a> ) . . . . .	42
4.1	Cronograma dos diferenciais de preço da gasolina (dobras/mês) . . . . .	54
4.2	Cronograma dos diferenciais de preço do gásóleo (dobras/mês) . . . . .	55
4.3	Cronograma dos diferenciais de preço do petróleo (dobras/mês) . . . . .	55
4.4	Decomposição dos diferenciais de preço da gasolina . . . . .	56
4.5	Decomposição dos diferenciais de preço do gásóleo . . . . .	56
4.6	Decomposição dos diferenciais de preço do petróleo . . . . .	56
4.7	Gráfico dos diferenciais de preço da gasolina (dobras/mês) por ano . . . . .	57
4.8	Gráfico dos diferenciais de preço do gásóleo (dobras/mês) por ano . . . . .	57
4.9	Gráfico dos diferenciais de preço do petróleo (dobras/mês) por ano . . . . .	57
4.10	Análise descritiva dos diferenciais de preço da gasolina, gásóleo e petróleo, por ano . . . . .	58
4.11	FAC e FACP dos diferenciais de preço da gasolina (gats) . . . . .	59
4.12	Série transformada dos diferenciais de preço da gasolina (diff(gats)) . . . . .	59
4.13	Resíduos dos diferenciais de preço da gasolina (modelo SARIMA(0,1,0) × (0,0,1) <sub>12</sub> ) . . . . .	60
4.14	Ajustamento do modelo SARIMA(0,1,0) × (0,0,1) <sub>12</sub> aos diferenciais de preço da gasolina. Dados originais (a preto) e ajustados (a verde) . . . . .	61
4.15	Previsão dos diferenciais de preço da gasolina para 2018-2019 (a azul) . . . . .	61
4.16	Dados reais (gats.real) e previsão da gasolina para o ano de 2017 modelo SARIMA(0,1,0) × (0,0,1) <sub>12</sub> . . . . .	62
4.17	Cálculo dos erros de previsão da gasolina . . . . .	62
4.18	FAC e FACP dos diferenciais de preço do gásóleo até 2016 . . . . .	63
4.19	Série dos diferenciais de preço do gásóleo transformada . . . . .	63
4.20	Resíduos do diferencial de preço do gásóleo modelo ARIMA(0,1,0) . . . . .	64
4.21	Teste Ljung-Box realizado a série do gásóleo modelo ARIMA(0,1,0) 2016 . . . . .	64
4.22	Gráfico comparativo do real e previsão 2017 de gásóleo . . . . .	66

4.23 Ajustamento do modelo $ARIMA(0, 1, 0)$ aos diferenciais de preço do gasóleo. Dados originais (a preto) e ajustados (a vermelha)	66
4.24 Previsão dos diferenciais de preço da gasóleo para 2018-2019 modelo $ARIMA(0, 1, 0)$	67
4.25 Tabela comparativa da previsão e real 2017 do Gasóleo	67
4.26 Cálculo dos erros de previsão do gasóleo	67
4.27 FAC e FACP dos diferenciais de preço do petróleo até 2016	68
4.28 Série dos diferenciais de preço do petróleo transformada	68
4.29 Resíduos do diferencial de preço do petróleo modelo $ARIMA(0,1,0)$	69
4.30 Teste Ljung-Box realizado a série do petróleo modelo $ARIMA(0,1,0)$ 2016	69
4.31 Gráfico comparativo do real e previsão 2017 de petróleo do modelo $ARIMA(0, 1, 0)$	70
4.32 Ajustamento do modelo $ARIMA(0, 1, 0)$ aos diferenciais de preço da gasolina. Dados originais (a preto) e ajustados (a azul)	71
4.33 Previsão dos diferenciais de preço do petróleo para 2018-2019 modelo $ARIMA(0, 1, 0)$	71
4.34 Tabela comparativa da previsão e real 2017 do petróleo	72
4.35 Beneficiários da política do diferencial de preço	72
6.1 Gráfico da série original da gasolina (Diferencial de preço, preços real e de venda)	77
6.2 Gráfico da série original do gasóleo (Diferencial de preço, preços real e de venda)	77
6.3 Gráfico da série original do petróleo (Diferencial de preço, preços real e de venda)	77
6.4 Teste Li-McLeod para a série do petróleo modelo $ARIMA(0, 1, 0)$	78
6.5 Teste Li-McLeod para a série do petróleo modelo $ARIMA(1, 1, 0)$	78
6.6 Teste Ljung-Box para a série do petróleo modelo $ARIMA(1, 1, 0)$	78
6.7 Teste Li-McLeod para a série do gasóleo modelo $(0, 1, 0)$	79
6.8 Teste Li-McLeod para a série do gasóleo modelo $(1, 1, 0)$	79
6.9 Teste Ljung-Box para a série do gasóleo modelo $(1, 1, 0)$	79

# Lista de Tabelas

- 4.1 Critérios de comparação relativos ao ajustamento dos diferenciais de preço da gasolina 61
- 4.2 Critérios de comparação relativos ao ajustamento dos diferenciais de preço do gasóleo 65
- 4.3 Critérios de comparação relativos ao ajustamento dos diferenciais de preço do petróleo 70





# 1

## Introdução

---

<b>1.1</b>	<b>Motivação do estudo</b>	<b>18</b>
<b>1.2</b>	<b>Breve história das Séries Temporais e alguns exemplos de aplicação</b>	<b>19</b>
<b>1.3</b>	<b>Organização da tese</b>	<b>20</b>

---

## 1.1 Motivação do estudo

Já no século XXI, e depois de muito se falar sobre a necessidade de se adoptarem energias alternativas que comportem menos riscos para o planeta, o petróleo (que já foi a causa, e ainda continua a ser, de muitas guerras) continua a ser a principal fonte de rendimento de muitos países.

Em geral, o preço dos produtos é estabelecido pela lei da oferta e da procura. Contudo, no que diz respeito ao petróleo o seu preço está em constante flutuação uma vez que a definição do seu valor rege-se por questões de geopolítica.

As flutuações constantes no preço do petróleo causa grandes incertezas na economia mundial e em São Tomé e Príncipe (STP), um país subdesenvolvido e com uma economia muito frágil e pequena, os efeitos negativos dessas flutuações constitui um grande problema para os seus governantes.

A prospecção de petróleo em STP teve seu início em 1997 e, através das avaliações feitas na época, tudo previa que a produção se iniciasse em 2002. Contudo, as perfurações de explorações feitas posteriormente produziram resultados insatisfatórios, tendo em conta que as reservas de hidrocarbonetos eram pequenas demais e o custo de extração tão grande para justificar novos investimentos, isto fez com que as empresas de grandes portes (Exxon, Chevron, Sinopec e Total) desistissem de explorar a Zona de Exploração Conjunta (ZEC), segundo o relatório do FMI n.º 14/2.

Tendo em conta a localização geográfica dos bolcos petrolíferos, criou-se uma ZEC com a Nigéria em 2001 e uma Zona de Exploração Exclusiva (ZEE).

São Tomé e Príncipe ainda não faz parte dos países produtores de petróleo, mas existem perspectivas para se realizar a primeira perfuração na ZEE em 2019.

Todo o combustível consumido é importado de Angola, o armazenamento do mesmo é assegurado pela Empresa Nacional de Combustível e Óleos (ENCO, SARL), cujo o acionista maioritário é a Sonangol. A ENCO, SARL, é a única importadora de combustível do país e, em conjunto com outras empresas, realiza a sua distribuição e venda.

A política de preços de combustível adaptada pelo governo de STP, antes do início desta tese, consistia em alterar menos vezes que possível os preços de venda ao público, independentemente das flutuações do preço a nível internacional, mesmo estando sujeito aos diferenças de preços (positivos ou negativos) que esta política implica. Este procedimento está sendo praticado há vários anos por sucessivos governos. Isto leva-nos a questionar se esta política é a mais adequada e se não existirá uma política melhor.

Para responder a esta questão iremos analisar o comportamento dos dados históricos do diferencial dos preços dos combustíveis (do gasóleo, da gasolina e do jet-A1/petróleo) desde 2001-2017, as causas que possam ter contribuído para os mesmos, descrever a evolução desses diferenciais no período em estudo e prever os valores para o ano de 2019.

Com a análise dos dados históricos, também iremos analisar qual o maior beneficiário da política de controlo do preço de combustível, se é o Estado santomense ou o consumidor final.

Tal como foi referido, da necessidade de um melhoramento no planeamento da política de compra de combustíveis por parte do governo de STP surgiu o interesse em estudar o diferencial dos preços dos combustíveis ao longo do período de estudo, nomeadamente através de modelos estocásticos que, ao longo do tempo, considerassem os registos estatisticamente relacionados entre si. Estes modelos estocásticos designam-se por modelos de Séries Temporais.

A vantagem em usar modelos estocásticos, é que é possível fazer previsões a longo prazo dos diferenciais dos combustíveis através dos modelos ajustados. As previsões a longo prazo ajudam na melhoria da planificação orçamental do governo, bem como numa melhor gestão da empresa ENCO, SARL.

## 1.2 Breve história das Séries Temporais e alguns exemplos de aplicação

### 1.2.1 Breve história das Séries Temporais

Desde há centenas de anos que são registadas quantidades, relativas a certos fenómenos, que variam ao longo do tempo. Através destes registos pretendeu-se sempre compreender a evolução, ou o comportamento, destes fenómenos. Outro dos grandes objetivos foi, através da compreensão desse comportamento, fazer previsões. As previsões são, em geral, entendidas como sendo para a valores futuros. Contudo, também se pode fazer previsões para valores passados (num certo espaço de tempo), que por algum motivo não foram registados.

A área da Estatística que estuda fenómenos que variam ao longo do tempo chama-se Séries Temporais. A área das Séries Temporais têm aplicação em diversas áreas que vão desde a Economia e Gestão, à Meteorologia, à Sociologia, às diversas áreas da Engenharia, às Ciências da Saúde, entre muitas outras.

*Os desenvolvimentos teóricos na análise de séries temporais começaram cedo com a teoria dos Processos Estocásticos. A primeira aplicação real de modelos autorregressivos (AR) a dados temporais vem do trabalho de Yule e Walker nas décadas de 1920 e 1930 (Stockholm University, Department of Statistics, 2017).*

Mais tarde, em 1937, Slutsky introduziu os modelos de médias móveis (MA) e, em 1938, Wold combinou ambos os modelos AR e MA e desenvolveu os modelos autoregressivos de médias móveis (ARMA). Os modelos ARMA são os modelos mais usuais para a modelação de séries temporais estacionárias. Contudo, Wold não conseguiu deduzir a função de verosimilhança do modelo ARMA e, portanto, não foi possível estimar-se os parâmetros do modelo, pelo que a aplicação destes a casos reais só foi possível em meados dos anos de 1960, quando os computadores foram capazes de realizar os cálculos necessários para otimizar os cálculos para a estimação dos parâmetros do modelo (Makridakis, S. and Hison, M., 1995).

Mais tarde, em 1970, os modelos ARMA tornaram-se populares através do trabalho de Box e Jenkins e, em 1976, estes autores publicaram um livro intitulado "Time Series Analysis: Forecasting and Control", onde sistematizaram a metodologia de análise de uma série temporal através dos seguintes passos: identificação do modelo, estimação dos seus parâmetros e sua validação. Box e Jenkins foram mais longe e introduziram também metodologias para transformar uma série não-estacionária numa série estacionária, introduzindo, assim, os modelos autoregressivos integrados de médias móveis sazonais (ARIMA), apropriados para a modelação de séries não estacionárias que apresentem tendência e/ou sazonalidade. (Time Series and Forecasting: Brief History and Future Research. Journal of the American Statistical Association, Vol. 95, No. 450 (Jun., 2000), pp. 638-643).

Muitos autores designam os modelos autoregressivos integrados de médias móveis sazonais simplesmente por ARIMA ou por ARIMA sazonais. Aqui iremos usar a designação também bastante usual de modelos SARIMA.

### 1.2.2 Alguns exemplos de aplicação

Tal como foi referido, existe uma longa aplicação dos modelos de Séries Temporais às mais variadas áreas. Iremos apresentar alguns trabalhos apenas os que dizem respeito à área dos combustíveis.

Tularam e Saeed (2016) aplicaram modelos ARIMA (entre outros modelos de Séries Temporais que não iremos abordar, como são o caso do método de Holt-Winters e da suavização exponencial) aos preços do petróleo nos Estados Unidos da América, entre 2011 e 2016, com o objectivo de com a sua aplicação, e comparação entre os vários métodos, poder indicar quer a quem elabora as políticas do sector, quer os estratégias de *marketing* qual o melhor método de previsão para o mercado do petróleo. Dos três métodos estudados os modelos ARIMA foi aquele que obteve os melhores resultados.

Outra aplicação dos modelos ARIMA pode encontra-se num artigo sobre a exportação de petróleo bruto da Nigéria, entre de 2002 e 2011 (Samson, B. and Mercy, O., 2015).

Muitos outros trabalhos têm surgido nesta área de aplicação, alguns mantendo esta metodologia ARIMA/SARIMA e outros usando outras metodologias mais avançadas que permitem incorporar

as grandes mudanças nas séries dos preços do petróleo causadas pelas grandes oscilações do mercado. Em particular, destaca-se a aplicação dos modelos vetorial auto-regressivos (VAR), por exemplo no trabalho de Kilian e Park (2009).

### 1.3 Organização da tese

De forma à alcançar os objectivos anteriormente definidos, está dissertação encontra estruturada nos cinco capítulos que são descritos em seguida.

Neste capítulo inicial descreve-se a motivação do estudo e apresenta-se uma breve síntese da literatura sobre a aplicação de modelos de séries temporais à produção e/ou consumos de combustíveis, como o caso do petróleo, da gasolina ou do gasóleo.

O segundo capítulo descreve-se o início os mercados de petróleo ao nível global, africano e nacional, incluindo uma breve explicação da cadeia petrolífera em STP e dos principais agentes que nela actuam. Destaque natural para a empresa ENCO, SARL, e a estrutura de preço de combustível que é elaborada pela mesma de acordo com as leis vigentes no país. Também se ilustra como é elaborado do mapa de estrutura de preços, onde se encontra o "diferencial de preço" que é o problema principal a estudar nesta dissertação.

No terceiro capítulo faz-se uma breve introdução à teoria das séries temporais, apresentando os conceitos mais relevantes e os modelos que se irá aplicar, de uma forma simples, mas tentando não perder o rigor matemático.

Nos fundamentos teóricos será dada especial atenção aos métodos quantitativos, nomeadamente ao método estocástico, já que no caso em estudo são-nos disponibilizados os dados históricos referentes aos anos anteriores, e os mesmos têm origem em eventos aleatórios, o que possibilita o desenvolvimento de um modelo com base neste tipo de métodos.

No quarto capítulo, após a identificação e a definição do problema, procede-se à recolha e análise de dados relativos ao diferencial de preços dos combustíveis (petróleo, gasolina e gasóleo). Na análise dos dados, as representações gráficas e as estatísticas resumo são duas etapas bastante importantes antes da modelação e previsão dos diferenciais dos preços dos combustíveis. Daí, numa primeira fase ter-se procedido a uma análise descritiva dos dados. Seguidamente, através da análise gráfica identificaram-se padrões e/ou tendências que caracterizam os nossos dados possibilitando, assim, a escolha dos modelos que melhor se adaptam a esses padrões. Por último, através dos modelos ajustados, procedeu-se à previsão de valores futuros dos diferenciais dos preços.

No quinto capítulo, apresenta-se as conclusões deste estudo. Ou seja, as vantagens da previsão do diferencial de preço do combustíveis a médio prazo, bem como a proposta de políticas alternativas e também identificar os beneficiários da política de fixação do preço praticado pelo Estado santomense durante os anos de 2001 a 2017. São também identificados aspectos a abordar em trabalhos futuros.

Todas as análises estatísticas foram efectuadas através do *software* R version 3.3.2 ([www.rproject.org](http://www.rproject.org)).

# 2

## Contextualização do caso de estudo

---

<b>2.1</b>	<b>Introdução</b>	<b>22</b>
<b>2.2</b>	<b>Caracterização da ENCO, SARL</b>	<b>22</b>
<b>2.3</b>	<b>Mapa de estrutura de preço</b>	<b>23</b>
<b>2.4</b>	<b>Mercado dos combustíveis</b>	<b>23</b>

---

## 2.1 Introdução

Este capítulo dá a conhecer a Empresa Nacional de Combustíveis e Óleo (ENCO, SARL) e o mercado dos combustíveis petrolíferos, descrevendo de forma detalhada as características das áreas onde se insere o caso de estudo a analisar ao longo desta dissertação. Para a compreensão do sector dos combustíveis torna-se fulcral a identificação das tendências de mercado, da dependência de cada mercado em termos energéticos e dos principais efeitos que as flutuações dos preços de combustíveis poderão causar nas economias, que aqui serão divididas e descritas a partir de uma escala decrescente (desde o nível mundial até ao nacional).

Para o mercado nacional descreve-se a situação actual da ENCO, SARL, no que diz respeito à sua política de importação de combustíveis. Dar-se-á especial atenção à descrição do centro de recepção e distribuição da ENCO, SARL (instalação de Neves, capital do distrito de Lembá, situada na zona norte da Ilha de São Tomé), especificando os tipos de combustíveis importados e armazenados, tal como as condições de importação e como se elabora o mapa de estrutura de preços. Será também descrito nesta dissertação todo o ciclo de distribuição e venda do petróleo em São Tomé e Príncipe e os principais responsáveis em cada etapa do ciclo, desde a sua requisição até à sua comercialização nos postos de abastecimento ao cliente final.

## 2.2 Caracterização da ENCO, SARL

A Empresa Nacional de Combustíveis e Óleos foi constituída sobre a forma de sociedade anónima de responsabilidade limitada, através do Decreto N.º 60/97 de 15 de Dezembro, com a denominação de ENCO, SARL. A sua sede é na Rua da Guiné, cidade de São Tomé, e tem como objecto social a importação e a comercialização de produtos petrolíferos, nomeadamente, combustíveis, óleos e massas lubrificantes e gás butano (plano de negócios 2018/2020).

Actualmente a estrutura accionista da Empresa é a seguinte:

SÓCIOS	ACÇÕES	%	U.S. DÓLARES
SONANGOL-HOLDINGS, LDA	7.756	77,56%	3.102.400
ESTADO SANTOMENSE	1.600	16%	640.000
PRIVADOS NACIONAIS	615	6,15%	246.000
TRABALHADORES DA ENCO, SARL	29	0,29%	11.600
<b>TOTAIS</b>	<b>10.000</b>	<b>100%</b>	<b>4.000.000</b>

Figura 2.1: Estrutura dos acionistas (Relatório e conta da ENCO, SARL 2017)

### Quota de mercado

A ENCO, SARL é a única empresa a operar no sector de importação de combustíveis brancos em São Tomé e Príncipe e controla maioritariamente o mercado de óleos e massas lubrificantes. Ao nível do mercado de gás a ENCO, SARL estabeleceu como política a expansão e o consumo, tendo em vista a conquista duma maior quota de mercado do GPL (gás de petróleo liquefeito).

### Formulação das estratégias

Na formulação de estratégia, a ENCO, SARL decidiu dividir o foco em três partes, que são a visão, missão e objectivo, como fio condutor para alcançar as suas metas em São Tomé e Príncipe (plano de negócios 2018/2020).

## Visão

- Ser uma empresa moderna, rentável e a maior em volume de negócios (comercialização de produtos petrolíferos e derivados);
- Desenvolver parcerias e competências para se posicionar como uma das líderes da indústria de hidrocarbonetos e petroquímica;
- Assumir em cada projeto uma postura de liderança na defesa do meio ambiente, em parceria com o Estado e a comunidade, rumo ao desenvolvimento sustentável;
- Transformação em empresa regional de energia digital e redução da forte dependência do estado santomense, EMAE e a Sonangol.

## Missão

- Criar valor para o acionista;
- Satisfazer o cliente;
- Contribuir para o bem-estar social dos colaboradores.

## Objectivos

Objetivo primário da ENCO, SARL é comercializar os serviços/produtos aos clientes de forma a que haja da parte destes o reconhecimento de que adquiriram valor, isto é, atendimento personalizado no qual se dá primazia indiscutível a qualidade, rapidez e eficiência (plano de negócios 2018/2020).

## 2.3 Mapa de estrutura de preço

O Mapa de estrutura de preço (MEP) é elaborado pela ENCO, SARL de acordo com a lei de importação e venda de combustíveis vigente no país, aquando da chegada do navio de combustível, ou seja, sempre que há uma importação elabora-se um MEP para cada produto.

Primeiramente lança-se a taxa de câmbio do dólar (USD) em relação à dobra do dia em que se está a elaborar o MEP, seguidamente coloca-se o preço CIF (USD), que multiplicado pela taxa de câmbio obtém-se o preço CIF em dobrar. Após essa passagem do preço CIF dólar para dobra, imputa-se as Receitas Fiscais e Custos Adicionais, para alcançar o preço base.

Ao preço base, acrescido da margem grossista, resulta o preço de venda grossista que somando os encargos com a distribuição e margem para o retalhista dá o preço real. **O diferencial de preço é a diferença entre o preço real e o preço de venda actual.** Ou seja, se o preço real for maior que o preço de venda actual o diferencial será negativo, caso contrário será positivo. Portanto, quando o diferencial de preço for negativo debita-se ao estado Santomense, e quando for positivo credita-se.

## 2.4 Mercado dos combustíveis

### 2.4.1 Mercado global

#### Flutuações de mercado

Segundo Ignacio Mártel <sup>1</sup>, nos anos 2001 a 2008, o preço do petróleo sofreu uma subida lenta durante esse período, devido diversos factores que contribuíram para essa tendência, como por exemplo: Os

<sup>1</sup>Professor catedrático de Electrónica na Universidad Complutense de Madrid e membro da Real Sociedade Espanhola de Física.

Designação	Percentagem (%)	Valor
Taxa de Câmbio (USD /Db.)		
Preço CIF (USD)		
CIF/ São Tomé		
<b>Receitas Fiscais</b>		
Direito de Importação	%	
Sobretaxa	%	
<b>Custos Adicionais</b>		
Fundo Rodoviário		
Emolumentos e Guias	%	
Agência do Despachante	%	
Encargos da ENCO		
<b>Preço Base</b>		
Margem para Grossista		
<b>Preço de venda Grossista</b>		
Encargos c/Distribuição		
Encargos c/Distribuição - Príncipe		
Margem para Retalhista		
<b>Preço Real</b>		
Diferença de Preços (PC-PR)		
<b>Preço de venda Actual</b>		

**Figura 2.2:** Mapa de Estrutura de Preço (ENCO, SARL)

atentados de 11 de setembro, a segunda Guerra do Golfo (2003), a eterna crise do Médio Oriente e a enorme bolha financeira que foi inchando desde os últimos anos do século XX, chegando próximo dos 140 dólares na primavera de 2008.

Após a primavera de 2008, começou a tão conhecida “Grande Recessão”, onde os preços sofreram grandes quedas, alcançando valores perto dos 40 dólares por barril, na primavera de 2009. Depois dos momentos difícil no mundo árabe, os preços vieram a estabilizar no ano 2012 e permaneceu assim até finais de 2014, onde os valores rondavam 80 a 100 dólares por barril.

Nos finais de 2014, começou a registar-se quedas sistemáticas nos preços e, tudo indica, que foi devido os seguintes factores: a redução da procura por parte da China, o aumento de produção por parte da Arábia Saudita e também o levantamento do embargo ao Irão após os acordos do verão de 2015 sobre o seu programa nuclear. Resumindo pode dizer-se que o preço por barril de petróleo é muito volátil, sofrendo ao longo de muitos anos, grandes subidas e descidas em curto período de tempo tornando, assim, difícil para qualquer governo orçamentar (ingresso ou custo) os valores dessa variável económica <https://blogs.publico.es/econonuestra/2016/02/13/la-historia-del-precio-del-petroleo-subidos-a-una-montana-rusa/>.

### **Movimentos do preço do petróleo bruto**

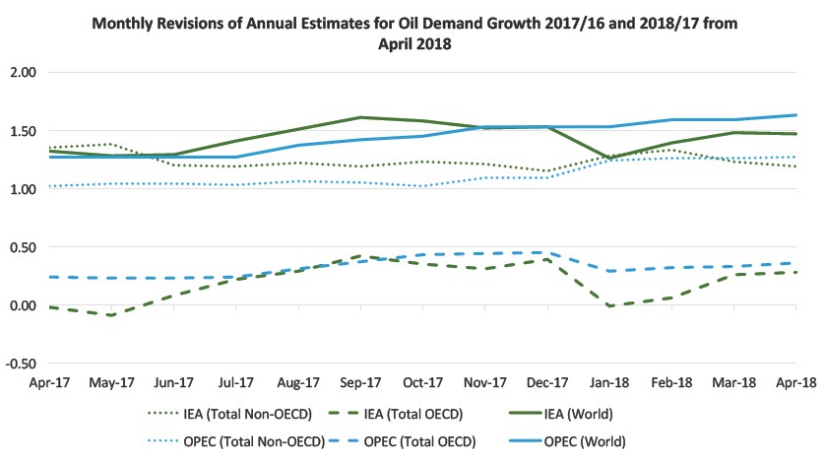
*Depois da redução repentina de preço por barril de petróleo a partir de 2014, tudo indicava uma redução nas quantidades produzidas, mas isto não se verificou, deixando prova de que não estávamos presentes num mercado de livre concorrência. O mercado de petróleo é do tipo oligopolista de competição, mas também são afectados pelas entradas de novos agentes e rupturas tecnológicas. Essas características, em nada ajuda à estabilização dos preços dessa matéria prima (Luis Dutra, 2017).*



## Procura mundial de petróleo

A organização dos países exportadores de petróleo (OPEP) projectou o crescimento da procura mundial de petróleo para 2018, em 1,63 milhões de barris diários (mb/d), esta previsão otimista é fruto do aumento inesperado da procura, principalmente dos países que fazem parte da Organização para a Cooperação e Desenvolvimento Económico (OCDE), e também apoiado pelo incremento das actividades industriais mais do que esperadas e pelas fortes actividades de mineração na América e na região da Ásia-Pacífico da OCDE (Relatório mensal da OPEP, Abril 2018).

Dentro da região não-OCDE, e alguns países da Ásia, a procura registou uma revisão em alta de 30 mil barris diários (tb/d) no primeiro trimestre de 2018 (1T18), resultante de uma revisão realizada que obteve resultados melhores do que o esperado. Em contrapartida, o crescimento da procura de petróleo foi ajustado para menos de 30 tb/d no 1T18 na região do Médio Oriente. Isso reflete principalmente desenvolvimentos da procura de petróleo regionais mais lentos do que o esperado. A procura mundial de petróleo para 2018 está prevista para uma média de 98,70 mb/d segundo a OPEP (Relatório mensal da OPEP, Abril 2018).



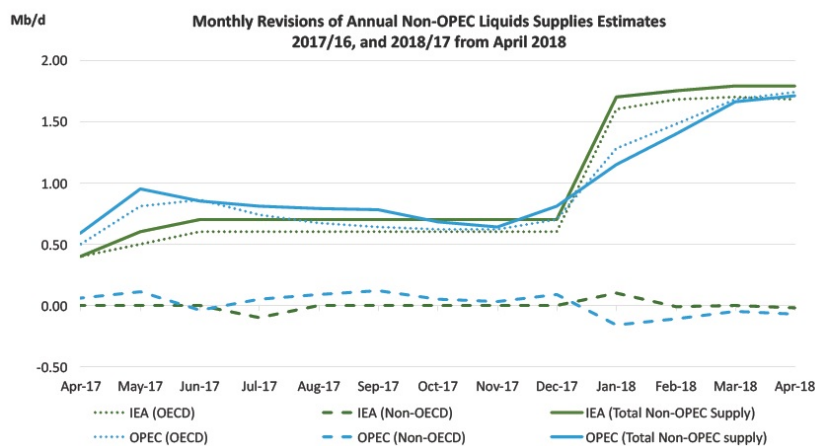
**Figura 2.3:** Análise comparativa, entre a IEF e a OPEP, sobre a procura de petróleo (Relatório mensal de Abril de 2018) ([https://www.ief.org/\\_resources/files/comparative-analysis/april-2018-comparative-analysis\\_final.pdf](https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf))

Análise comparativa do IEF dos Relatórios Mensais da AIE e da OPEP sobre o Mercado de Petróleo Abril de 2018

## Oferta mundial de petróleo

Em quatro revisões mensais consecutivas as projeções para o ano 2018, realizadas pela OPEP e pela Agência Internacional de Energia (AIE), as diferenças entre os dados diminuíram ao longo das revisões, estando a oferta não-OPEP em 59,61 mb/d e 59,93 mb/d, respectivamente. Isto representa um crescimento ano após ano de 1,71 mb/d e 1,79 mb/d (Forúm Internacional para a Energia, que se realizou em Riyadh, Arábia Saudita, Abril 2018, [https://www.ief.org/\\_resources/files/comparative-analysis/april-2018-comparative-analysis\\_final.pdf](https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf)).

Tanto a AIE, como OPEP, apresentam dados que indicam uma diminuição da produção de petróleo bruto da OPEP em mês de Março de 2018. Esta diminuição continua mês após mês, segundo fontes secundárias da OPEP, tem apresentando diminuição de 0,2 mb/d para uma média de 31,83 mb/d de acordo com os dados da AIE. Segundo os dados da OPEP, a oferta de petróleo bruto da OPEP-14 também diminuiu 0,2 mb/d em relação ao mês anterior, para uma média de 31,96 mb/d (Forúm Internacional para a Energia, que se realizou em Riyadh, Arábia Saudita, Abril 2018, em [https://www.ief.org/\\_resources/files/comparative-analysis/april-2018-comparative-analysis\\_final.pdf](https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf)).



**Figura 2.4:** Análise comparativa, entre a IEF e a OPEP, sobre a oferta de petróleo (Relatório mensal de Abril de 2018) ([https://www.ief.org/\\_resources/files/comparative-analysis/april-2018-comparative-analysis\\_final.pdf](https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf))

### Análise global

A procura global por petróleo tem apresentado solidez no seu crescimento, estando os valores projetados para o ano 2018, a rondar um aumento de 3,8% e 3,9%, de acordo com os relatórios da OPEP e da AIE de Março.

Os dados globais apresentam certa contradição entre os parâmetros de crescimento económico e indicadores de liderança e sentido, visto que o primeiro aponta um valor sempre otimista, enquanto que o segundo apresenta alguma moderação na dinâmica de crescimento. Esse crescimento da procura poderá ser afectado pelo aumento da volatilidade nos mercados e tensões disruptivas no comércio global. Isto pode servir de motivação para fugir da era de flexibilização quantitativa, que poderá influenciar o crescimento económico (Forum Internacional para a Energia, que se realizou em Riyadh, Arábia Saudita, Abril 2018, [https://www.ief.org/\\_resources/files/comparative-analysis/april-2018-comparative-analysis\\_final.pdf](https://www.ief.org/_resources/files/comparative-analysis/april-2018-comparative-analysis_final.pdf)).

### Mercado africano

Após a abordagem do mercado global de combustíveis, passar-se-á a analisar o mercado africano, onde se vai relatar algumas das suas características.

África é o terceiro maior continente, com 54 países numa extensão de 30.200 mil quilómetros quadrados, e um bilião de pessoas. A sua produção de petróleo representa um décimo da produção mundial. A China é um dos principais parceiros dos países africanos e essa parceria também passa pelo sector do petróleo, onde um terço das importações de petróleo é garantida pelos países africanos (Dr Thomas Chaize, 2010, <http://www.dani2989.com/matiere1/chinaafrica0810pt.htm>).

Os governos africanos têm total autonomia na tomada de decisão no sector petrolíferos, tanto na escolha das companhias internacional que desejam que operem nas suas fronteiras, como também na aplicação das receitas proveniente do petróleo. As instituições governamentais africanas, principalmente do Golfo da Guiné, carecem de estruturas (recursos humanos) com experiencia na negociações de contratos de petróleo. Tendo em conta a quantidade de companhias estrangeiras que operam em África, podemos considerar este mercado bastantes aberto (Gary e Karl, 2003).

A OPEP está constituída neste momento por 15 países membros, sendo que 8 deles são dos 15 maiores produtores mundiais. A África está representada na OPEP pelos seguintes países: Angola, Argélia, Gabão, Guiné Equatorial, Líbia, Nigéria e República do Congo. Onde podemos destacar a Nigéria e Angola pela luta ao pódio de maior produtor de petróleo da África [https://www.opec.org/opec\\_web/en/about\\_us/25.htm](https://www.opec.org/opec_web/en/about_us/25.htm)

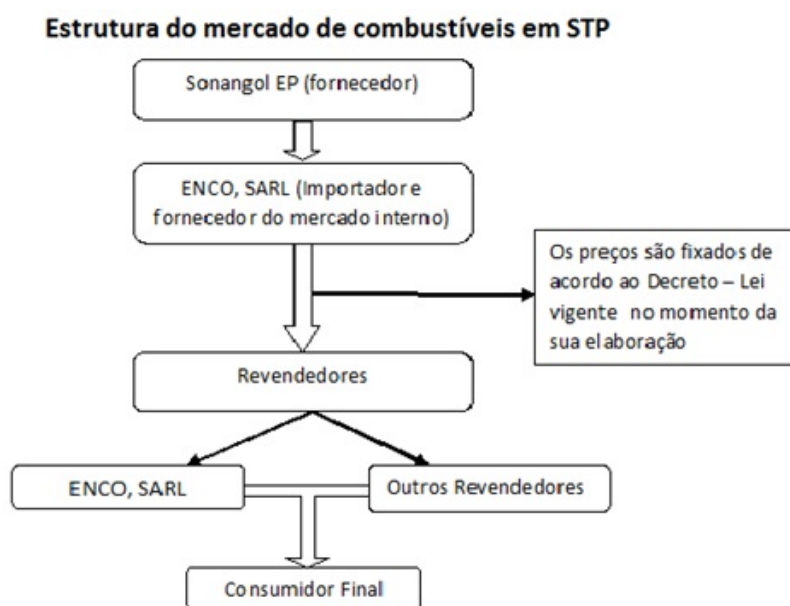
## Melhores perspectivas e desenvolvimentos

Na África ocidental, Angola e Nigéria são os maiores exportadores de petróleo e ambos estão empenhados em arrancar com novos projetos de petróleo e gás em 2018 como forma de estimular a indústria. As características do continente africano são propícias para a implementação de energias renováveis. Por esse motivo, existe um aumento constante de projetos que visam aproveitar esta abundância em recursos naturais para se materializar ainda mais a utilização desse tipo de energia, porque a produção de energia contínua é um dos primeiros passos para o progresso. (africaHRsolutions em <http://africa-hr.com/pt/2017/06/20/petroleo-e-gas-a-africa-em-busca-de-novos-projetos/>)

### 2.4.2 Mercado nacional

#### O ciclo do petróleo em São Tomé e Príncipe

A indústria petrolífera nacional está intimamente ligada à publicação da designada “Lei Quadro da Actividade Petrolífera”, Lei n.º 4/2000. Contudo, as taxas e os preços aplicados aos importadores e revendedores são regulados por Decreto Lei. São Tomé e Príncipe, ao longo de vários anos, continua a ter um só importador de combustíveis, que é a ENCO, SARL, mas existem diversos revendedores, incluindo também a ENCO, SARL, visto que a mesma também possui postos de abastecimento de venda para o consumidor final. Os produtos são expedidos na instalação da ENCO, SARL, em Neves e, em seguida, são distribuídos através dos camiões cisterna para os revendedores, onde os preços praticados pelos mesmos devem ser iguais, de acordo com o Decreto lei vigente.



**Figura 2.5:** Mapa de estrutura de mercado (Hearton De Menezes, 2017)

O transporte desta matéria-prima, sendo São Tomé e Príncipe um país que importa a totalidade do petróleo de que necessita, é efectuada pelos navios pertencentes à Sonangol EP, que é o único fornecedor de combustíveis da ENCO, SARL e na conjuntura actual do país.

Na armazenagem dos produtos, são utilizados os tanques localizados na instalação de Neves, que possui um parque de armazenamento e enchimento com 10 tanques e com uma capacidade de armazenagem de 22.890.000 litros, sendo 15.100.000 litros para gasóleo, 3.290.000 litros para gasolina e os restantes para jet-A1/petróleo, com um consumo médio diário de 110.000 litros de gasóleo, 30.000 litros de gasolina e 30.000 litros também para jet-A1/petróleo.

Quanto à distribuição de produtos petrolíferos em São Tomé e Príncipe, esta é efectuada por camiões-cisterna estando a sua tecnologia enquadrada pela regulamentação do transporte rodoviário de substâncias perigosas. Este aspecto realça a elevada dependência do transporte

rodoviário por parte da actividade de distribuição de produtos petrolíferos. As empresas distribuidoras cobram uma taxa (regulamentada pelo estado) por cada litro de combustível transportado. Quanto à distribuição de Jet-A1, é transportado da Instalação de Neves para a Instalação do Aeroporto internacional, através dos caminhões cisterna (pertencentes à ENCO, SARL) próprio para transporte desse produto.

A venda é feita nos postos de abastecimento pertencentes aos revendedores espalhados pelo país, onde o preço praticado é determinado pelo Estado. A Ilha do Príncipe não foge a regra, onde o custo transporte marítimo é assumido pela ENCO, SARL, para que os preços praticados pelos revendedores da Ilha do Príncipe sejam iguais ao resto do país. Quanto à venda de Jet-A1, é feita pela ENCO, SARL através da Instalação que a mesma possui no aeroporto internacional de São Tomé e Príncipe, onde o preço praticado não tem interferência do Estado santomense.

### **Efeitos do câmbio**

O momento de queda no preço do petróleo leva alguns exportadores a se prepararem para uma significativa queda de receitas. Ao mesmo tempo, países importadores como é o caso de São Tomé e Príncipe podem beneficiar, já que o preço tende a cair para os seus consumidores.

São Tomé e Príncipe, não obstante as flutuações dos preços dos combustíveis, também é afetado pelo câmbio do dólar em relação a dobra, visto que as compras são feitas em dólar e as vendas feitas em dobra. Na elaboração do Mapa de Estrutura de Preço (MEP), quando se realiza a passagem do preço CIF (dólar) para preço CIF (dobra), materializa-se o efeito do câmbio sobre o preço de venda real e, conseqüentemente, ao diferencial de preço. Não obstante os efeitos acima citados, no processo de pagamento ao fornecedor também se verifica o efeitos do câmbio, tendo em conta que a empresa importadora (Enco, SARL) necessita comprar dólar para liquidar as facturas ao fornecedor.

### **Efeitos sobre a inflação**

Na economia mundial, os preços da maioria dos produtos derivam do preço de petróleo e, em São Tomé e Príncipe, essa dependência é ainda maior, visto que grande parte dos produtos que se consomem são importados, a produção de energia que alimenta as casas é quase na totalidade produzida em centrais térmicas.

Os canais através dos quais os preços do petróleo afectam os preços do consumidor podem dividir-se em efeitos de primeira ordem - desagregados ainda em efeitos directos e indirectos - e em efeitos de segunda ordem (Esteves e Neves, 2004).

Os efeitos de primeira ordem, como citado anteriormente, podem ser directos ou indirectos. Ambos os casos espelham que tanto os combustíveis, como os bens e serviços que dependem destes para o seu funcionamento, estão incluídos no índice de preços do consumidor (IPC)<sup>2</sup>. A diferença é que no efeitos directos o impacto incide sobre as componentes energéticas do índice de preços (taxas aplicadas à gasolina, gasóleo, gás, etc) e nos indirectos, o impacto incide sobre as componentes do índice que têm um conteúdo energético (preços de passagem de avião, barco, táxi, etc).

O impacto dos efeitos directos sobre o IPC depende da percentagem de produtos energéticos consumidos. Ou seja, quanto maior for o consumo de produtos energéticos, maior será o impacto directo sobre o IPC de uma determinada variação dos preços de energia.

A dimensão dos efeitos de primeira ordem directos também podem ser afectado pela importância e pela estrutura das taxas de impostos. Se a taxa do imposto for baseada em percentagem (por exemplo, 5% + o preço por unidade de produto) o aumento do custo dos produtos energéticos será totalmente transmitido ao consumidor. Entretanto, se for um valor fixo, independente (por exemplo, 5 dobras + o preço por unidade produto) o aumento do custo dos produtos energéticos será apenas parcialmente transmitido ao consumidor, porque o valor a pagar será sempre 5 dobras independentemente do preço de produto (Esteves e Neves, 2004).

---

<sup>2</sup>índice que mede o custo de um determinado cabaz fixo, e que pode variar de acordo com o país, de bens e serviços de consumo.

No caso de São Tomé e Príncipe, os impostos que recaem sobre os produtos petrolíferos são calculados com base nas percentagem (por exemplo, 5% + o preço por unidade de produto). Portanto, um aumento do custo dos produtos energéticos será totalmente transmitido ao consumidor.

Ao contrário dos efeitos directos, a magnitude de afectação dos efeitos indirectos é difícil de avaliar. Contudo, grande parte dos preços dos produtos estão afectados por esses efeitos. Caso o preço do combustível sofresse um aumento considerável, muitos dos preços dos bens e serviços com alguma dependência dos combustíveis reagiriam a esse aumento. Os exemplos mais claros em São Tomé e Príncipe são os serviços de táxi, e moto-táxi, que têm quase total dependência do combustível para o seu funcionamento. Entretanto, a magnitude da afectação depende da importância do combustível como factor de produção (Esteves e Neves, 2004).

No que diz respeito aos efeitos de segunda ordem, estes *estão associados com a reacção macroeconómica à variação dos preços do petróleo, os mesmos causam efeitos contemporâneos e actuam de forma desfasados sobre as restantes componentes do IPC* (Esteves e Neves, 2004).

Esses efeitos são o fruto de uma relação entre salários e preços. Sempre que houver um aumento do preço no consumidor, e os trabalhadores exerçam pressão sobre as entidades patronais para ajustar os salários nominais<sup>3</sup> de acordo com o aumento do preço no consumidor, em vez de aceitarem os salários reais<sup>4</sup> mais baixo, apareceriam outros tipos de pressões inflacionistas, através de uma espiral salários-preços. *Neste caso, a flexibilidade do mercado de trabalho e política monetária na formação das expectativas de inflação têm grande influência na dimensão desses efeitos no IPC* (Esteves e Neves, 2004).

---

<sup>3</sup>É o valor do trabalho expresso em moeda.

<sup>4</sup>É o que se pode comprar (poder de compra) com o salário nominal.

# 3

## Métodos quantitativos estocásticos de ajustamento e de previsão

---

3.1	Introdução . . . . .	31
3.2	Séries temporais: breve introdução . . . . .	31
3.3	Modelos para séries temporais lineares estacionárias . . . . .	36
3.4	Modelos para séries temporais lineares não estacionárias . . . . .	41
3.5	Modelação de séries temporais . . . . .	45
3.6	Previsão de séries temporais . . . . .	51

---

## 3.1 Introdução

Neste capítulo, iremos descrever o enquadramento teórico sobre as séries temporais, principalmente sobre os métodos de previsão. Existem dois tipos de métodos de previsão, o método qualitativo onde frequentemente não se utiliza nenhum modelo matemático formal, porque se houver dados disponível, normalmente eles não são considerados como representativos do futuro. Enquanto que o método quantitativo, existem sempre dados históricos das variáveis de interesse disponíveis, também examina as relações de causa e efeito da variável com outras variáveis relevantes, e normalmente se utiliza modelos matemáticos e software para processamentos dos dados, mas actualmente a tendência é usar os dois métodos de maneira complementar.

Os métodos quantitativos estão divididos em métodos determinísticos, que não contém nenhuma variável aleatória, e os métodos estocásticos, que contém pelo menos uma variável aleatória, e com ela vem associada a incerteza, que são características semelhantes ao caso de estudo desta dissertação, já que são-nos disponibilizados os dados históricos referentes aos anos anteriores, e os mesmos têm origem em eventos aleatórios, o que possibilita o desenvolvimento de um modelo com base neste tipo de métodos.

No decorrer deste capítulo, iremos fazer uma breve abordagem às séries temporais e aos respectivos métodos de previsão, mas somente os métodos quantitativos estocásticos, considerando que é o método utilizado nesta dissertação. Também se irá descrever os passos para modelação de séries temporais e, para finalizar, explicar como fazer a previsão a partir do modelo ajustado e quais são os critérios para se avaliar a qualidade dessas previsões.

## 3.2 Séries temporais: breve introdução

Uma série temporal (também designada por série cronológica) é uma sucessão de observações, associadas a um certo acontecimento aleatório, que são registadas ao longo do tempo e que estão estatisticamente relacionadas, ou seja, que são dependentes entre si (até um certo horizonte de tempo). Os casos mais usuais de séries temporais, sobre os quais nos vamos centrar, têm por pressuposto que as observações ao longo do tempo são igualmente espaçadas.

Analisar uma série temporal tem por detrás dois grandes objectivos: o primeiro consiste em conseguir identificar o tipo de evolução ao longo do tempo do acontecimento em estudo e traduzir essa evolução num modelo matemático. Ou, por outras palavras, modelar a dependência das observações. O segundo em fazer previsões de valores futuros do acontecimento em estudo, através desse modelo matemático.

As representações gráficas de séries temporais têm um papel fundamental e devem constituir o primeiro passo de qualquer análise pois através destas é possível identificar padrões e, deste modo, pensar em classes de modelos que se possam vir a adequar aos dados em estudo.

Existem diversos padrões, mas iremos apenas abordar os casos de séries que oscilam em torno de uma média fixa, de séries que apresentam uma evolução crescente (ou decrescente) e as que apresentam movimentos oscilatórios.

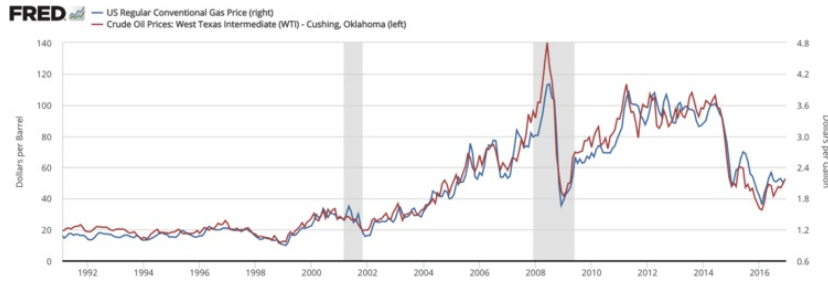
Quando uma série temporal apresenta uma evolução crescente (ou decrescente), ao longo do tempo diz-se que possui tendência crescente (ou decrescente). Quando apresenta movimentos oscilatórios (ou ciclos) que se repetem ao longo do tempo com o mesmo período (ou com a mesma regularidade) diz-se que possui sazonalidade. Ou seja, a série tem uma variação ao longo do tempo de natureza periódica. Se estes ciclos não tiverem a mesma regularidade ao longo do tempo, ou seja, não se repetirem num mesmo período fixo, então diz-se que a série tem um comportamento cíclico. Uma série temporal pode ainda apresentar conjuntamente tendência e sazonalidade, sazonalidade e ciclicidade ou tendência, sazonalidade e ciclicidade.

Os padrões sazonais surgem quando a sucessão das observações, num certo horizonte temporal, é influenciada por factores sazonais, por exemplo, o mês, o ano, o dia da semana, entre outros.

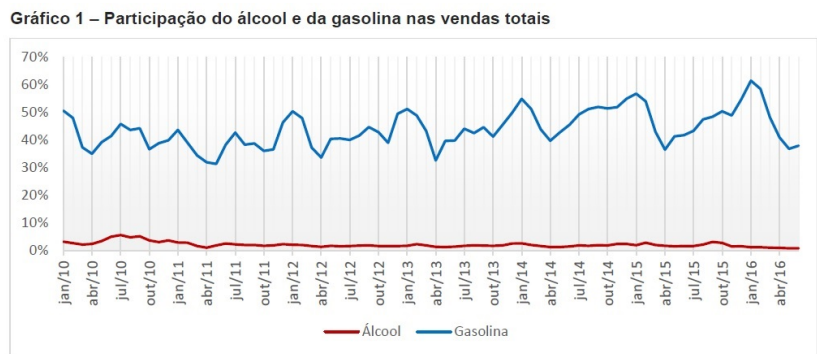
Se uma série que oscila em torno de uma média fixa diz-se que é *estacionária*. Caso possua tendência e/ou sazonalidade então diz-se que é *não estacionária*. Se tiver um comportamento cíclico, mas não tiver tendência nem sazonalidade, então também se diz *estacionária*. Em resumo, podemos dizer que, em geral, uma série estacionária não tem um padrão previsível a longo prazo.

Seguidamente apresentaremos alguns exemplos de séries temporais.

Tal como referido anteriormente, um dos objectivos da análise de séries temporais consiste



**Figura 3.1:** Série temporal com tendência crescente até 2008 (FRED government work  
<https://fred.stlouisfed.org/series/GASREGCOVM>)



**Figura 3.2:** Série temporal com sazonalidade (Samia, Keimara e Pedro, 2017  
<https://online.unisc.br/seer/index.php/cepe/article/download/9063/6870>)

em conseguir identificar o tipo de evolução ao longo do tempo do acontecimento em estudo e traduzir essa evolução num modelo matemático que melhor consiga descrever o comportamento do fenómeno em estudo, que evolui ao longo do tempo. Dizemos, portanto, que o fenómeno depende do tempo e do acaso.

O modelo matemático que descreve este tipo de fenómenos ao longo do tempo é denominado *processo estocástico*. Por este motivo vamos, na secção seguinte, apresentar algumas generalidades sobre processos estocásticos.

### 3.2.1 Conceitos gerais sobre processos estocásticos

Um processo estocástico  $\{X(t), t \in T\}$  (ou  $\{X_t\}_{t \in T}$ ) é uma sucessão de variáveis aleatórias indexadas no tempo (ou no espaço. Nesta abordagem apenas nos iremos referir ao tempo).

Define-se variável aleatória (v.a.) como uma função, digamos  $X(\omega)$ , que associa a cada acontecimento  $\omega \in \Omega$ , sendo  $\Omega$  o espaço amostral, um valor pertencente a  $\mathbb{R}$ . Usualmente usa-se apenas a letra  $X$  para designar a v.a. não esquecendo que o valor assumido  $X$  depende do ponto  $\omega$ . Tal como na definição de uma v.a., também um processo estocástico (p.e.) é uma função, digamos  $X(t, \omega)$ , que associa a cada acontecimento  $\omega \in \Omega$  e a cada ponto  $t \in T$  um valor em  $E$ , a que se chama *espaço dos estados do processo*. A  $T$  chama-se *espaço parâmetro*.

Se o tempo  $t$  for fixo  $X(t, \omega) \equiv X(\omega) \equiv X$  é uma v.a., e se  $\omega$  for fixo,  $X(t, \omega) \equiv X(t)$  chama-se uma *realização* ou *trajectória* do processo.

O tempo, ou conjunto qualquer de índices  $T$ , pode ser discreto ou contínuo e o espaço de estados  $E$  também pode ser discreto ou contínuo. Assim, os p.e. classificam-se do seguinte modo:

(i) Quanto à natureza do espaço parâmetro:

Se  $T$  é finito ou infinito numerável o p.e. diz-se de *parâmetro discreto*. Se, pelo contrário,  $T$  é infinito não-numerável o p.e. diz-se de *parâmetro contínuo* (ou *permanente*).

(ii) Quanto à natureza do espaço de estados:



Se  $E$  é finito ou infinito numerável o p.e. diz-se *discreto*, se  $E$  é infinito não-numerável o p.e. diz-se *contínuo*. Se  $E \subseteq \mathbb{Z}$  o p.e. diz-se de *valores inteiros*.

Uma outra classificação dos processos estocásticos diz respeito ao tipo de relação de dependência entre as suas variáveis. Ou seja, tal como referimos inicialmente, um processo (ou uma sua realização, isto é, uma série temporal) pode ser estacionária ou não estacionária. Dentro da estacionariedade ainda se pode definir estacionariedade forte ou estacionariedade fraca.

De agora em diante iremos apenas considerar  $T = \mathbb{Z}$ .

**Definição 3.2.1.** Um processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$  diz-se *fortemente estacionário* (ou *estacionário em sentido estrito*) se,  $\forall n \geq 1, \forall k \geq 0$  e qualquer escolha de índices  $(t_1, t_2, \dots, t_n) \in \mathbb{Z}$ , a distribuição dos vectores

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \quad \text{e} \quad (X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k})$$

é *idêntica*.

Como esta propriedade é muito difícil de provar na prática usualmente recorre-se a uma definição de estacionariedade mais fraca mas fácil de verificar, que apresentamos a seguir.

**Definição 3.2.2.** Um processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$  diz-se *estacionário de segunda ordem* (ou *fracamente estacionário* ou, ainda *estacionário em covariância*) se, para todo o  $t$ ,  $E[X_t^2] < +\infty$ ,  $\forall k \geq 0$  e

1.  $E[X_t] = m$  (constante),  $t \in \mathbb{Z}$
2.  $Var[X_t] = \sigma^2$  (constante),  $t \in \mathbb{Z}$
3.  $Cov(X_t, X_s) = Cov(X_{t+k}, X_{s+k}) = \gamma(|t-s|)$ ,  $t, s \in \mathbb{Z}$ .

Ou seja, a média e a variância são constantes ao longo do tempo e a covariância entre as variáveis  $X_t$  e  $X_s$  (representada por  $\gamma(t, s)$ ) depende apenas da distância temporal entre as variáveis,  $|t-s|$ .

Em geral, na teoria de séries temporais quando aparece referido estacionariedade, sem qualquer outro atributo, isto quer dizer que se está a referir à estacionariedade fraca.

O exemplo mais simples, e usual, de p.e. estacionário é o chamado *ruído branco* (do inglês *white noise*) e que não é mais do que um processo puramente aleatório e que irá servir para a definição de quase todos os modelos de séries temporais.

**Definição 3.2.3.** Um processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$  diz-se um *processo de ruído branco* se

1.  $E[\epsilon_t] = \mu_\epsilon$
2.  $Var[\epsilon_t] = \sigma_\epsilon^2$
3.  $Cov(\epsilon_t, \epsilon_{t+k}) = \gamma(k) = 0$ ,  $k = \pm 1, \pm 2, \dots$

Em geral considera-se  $\mu_\epsilon = 0$ . Assim, sempre que nos referirmos ao processo de ruído branco estamos a considerar que  $\mu_\epsilon = 0$ . O processo de ruído branco, de média nula e variância  $\sigma_\epsilon^2$  representa-se, abreviadamente, como  $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$ , ou, do inglês,  $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ .

Por exemplo, uma série temporal  $X_t$  poderia ser descrita ao longo de um certo período de tempo em termos de uma componente regular, digamos uma tendência  $\mu_t$ , mais o acaso (ou o *ruído*). Ou seja,

$$X_t = \mu_t + \epsilon_t, \quad t \in \mathbb{Z}$$

### 3.2.2 Funções de autocovariância, autocorrelação e autocorrelação parcial

Vamos considerar que  $\{X_t\}_{t \in \mathbb{Z}}$  é um p.e. estacionário. Então,  $\forall t \in \mathbb{Z}$ ,  $E[X_t] = m$ ,  $Var[X_t] = \sigma^2$  e, portanto, a *função de autocovariância*,  $\{\gamma(k), k \in \mathbb{Z}\}$ , é dada por

$$\gamma(k) = Cov(X_t, X_{t+k})$$

**Nota:**  $\gamma(0) = Var[X_t]$ .

Por este motivo, a *função de autocorrelação* (designada usualmente pela sigla FAC),  $\{\rho(k), k \in \mathbb{Z}\}$ , vem dada por

$$\rho(k) = Corr(X_t, X_{t+k}) = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var[X_t]Var[X_{t+k}]}}$$

ou seja,

$$\rho(k) = \frac{Cov(X_t, X_{t+k})}{Var[X_t]} = \frac{\gamma(k)}{\gamma(0)}$$

**Nota:**  $\rho(0) = 1$ .

Prova-se (Pacheco, 2001) que se  $\{X_t\}_{t \in \mathbb{Z}}$  é um p.e. estacionário então,  $\forall k \in \mathbb{Z}$ ,

- $\gamma(k) = \gamma(-k)$ ,
- $|\gamma(k)| \leq \gamma(0)$ ,
- $\rho(k) = \rho(-k)$ ,
- $|\rho(k)| \leq 1$ .

Chama-se *função de autocorrelação parcial de ordem k* (abreviadamente FACP),  $\{\phi(k, k), k \in \mathbb{Z}\}$ , à correlação entre as variáveis  $X_t$  e  $X_{t+k}$  sem que as variáveis intermédias ( $X_{t+1}, \dots, X_{t+k-1}$ ) tenham qualquer influência.

Assim, a FACP é definida como

$$\phi(k, k) = \frac{Cov(X_t, X_{t+k} | X_{t+1}, \dots, X_{t+k-1})}{\sqrt{Var[X_t | X_{t+1}, \dots, X_{t+k-1}]} \sqrt{Var[X_{t+k} | X_{t+1}, \dots, X_{t+k-1}]}}$$

Ao valor de  $k$  chama-se, por vezes, a *dimensão da memória*.

Prova-se ainda que se pode estabelecer uma relação entre as FAC e as FACP do seguinte modo (Pacheco, 2001)

$$\rho(h) = \phi(k, 1)\rho(h-1) + \phi(k, 2)\rho(h-2) + \dots + \phi(k, k)\rho(h-k), \quad h = 1, 2, \dots, k.$$

**Nota:**  $\phi(0, 0) = 1$  e  $\phi(1, 1) = \rho(1)$ .

**Nota:** Se  $\epsilon_t \sim RB(\mu_\epsilon, \sigma_\epsilon^2)$ , com  $\mu_\epsilon$  nulo ou não, então,

$$\gamma_\epsilon(k) = \begin{cases} \sigma^2, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

e

$$\rho_\epsilon(k) = \phi_\epsilon(k, k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

### 3.2.3 Estimação da média, variância e funções de autocovariância, autocorrelação e autocorrelação parcial

Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um p.e. estacionário. Para conhecermos o processo é necessário conhecer algumas medidas definidas anteriormente, tais como a média, a variância e as funções de autocovariância, autocorrelação e autocorrelação parcial. Como estas medidas são desconhecidas é preciso estimá-las.

Seja  $(X_1, X_2, \dots, X_n)$  uma realização do processo estacionário  $\{X_t\}_{t \in \mathbb{Z}}$ . O estimador do valor esperado do processo,  $E[X_t] = \mu$ , será dado por

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{t=1}^n X_t.$$

Relativamente à função de autocovariância,  $\gamma(k) = E[(X_t - \mu)(X_{t+k} - \mu)]$ , são considerados dois estimadores, mas o que apresenta melhores propriedades estatísticas é o dado pela seguinte expressão

$$\hat{\gamma}(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}).$$

Portanto, o estimador para a função de autocorrelação será dado por

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}, \quad k > 0.$$

O estimador da função de autocorrelação parcial é obtido a partir do estimador da função de autocorrelação, através de um método designado de *Durbin-Levinson*, tal que:  $\hat{\phi}(0,0) = 1$ ,  $\hat{\phi}(1,1) = \hat{\rho}(1)$  e

$$\hat{\phi}(k,k) = \frac{\hat{\rho}(k) - \sum_{j=1}^{k-1} \hat{\phi}(k-1,j)\hat{\rho}(k-j)}{1 - \sum_{j=1}^{k-1} \hat{\phi}(k-1,j)\hat{\rho}(j)}, \quad k > 1$$

e

$$\hat{\phi}(k,j) = \hat{\phi}(k-1,j) - \hat{\phi}(k,k)\hat{\phi}(k-1,k-j), \quad j = 1, \dots, k-1.$$

Todos estes estimadores verificam boas propriedades estatísticas (Pacheco, 2001).

**Nota:** Se  $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$ , então, para  $k \in \mathbb{N}$  e  $n$  suficientemente grande,

$$\hat{\rho}_\epsilon(k) \approx N\left(0, \frac{1}{n}\right)$$

e

$$\hat{\phi}_\epsilon(k,k) \approx N\left(0, \frac{1}{n}\right)$$

onde  $\approx$  significa que o estimador segue aproximadamente uma distribuição Normal.

Seguidamente iremos apresentar um exemplo de uma série de ruído branco com as respectivas FAC e FACP. Nos gráficos das FAC e FACP (aliás de todas, em quaisquer modelos) as linhas horizontais a tracejado representam os intervalos de confiança a 95% destas funções. Ou seja, os valores em cada  $k$  quer da FAC, quer da FACP, que estiverem fora dos intervalos de confiança são considerados diferentes de zero, enquanto os que estiverem dentro dos intervalos são considerados nulos.

Como se pode observar, quase todos os valores quer da FAC quer da FACP estão dentro dos intervalos de confiança. Existem dois que estão fora, mas pela definição de intervalo de confiança a 95%, por exemplo, se tivermos 100 valores 5 têm probabilidade de estar fora.

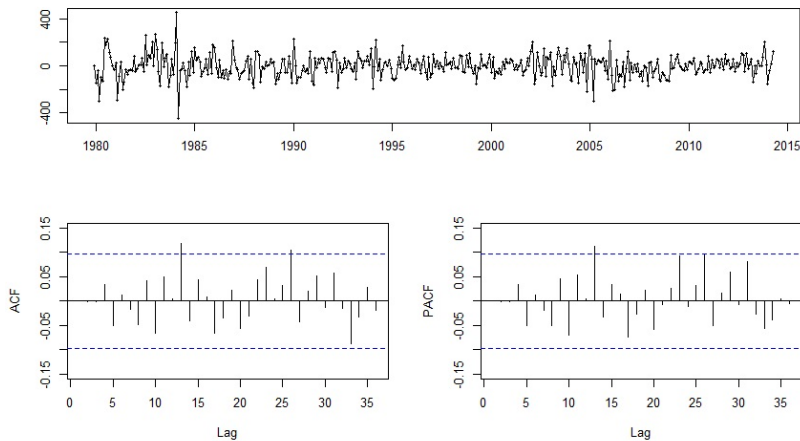


Figura 3.3: Série temporal, FAC e FACP de um ruído branco

### 3.3 Modelos para séries temporais lineares estacionárias

Iremos agora introduzir uma classe de modelos lineares de séries temporais estacionárias, designados por modelos autoregressivos e de médias móveis, abreviadamente modelos ARMA. Estes modelos são combinações de modelos mais simples, os chamados modelos autoregressivos (AR) e de modelos de médias móveis (MA).

Em geral, os modelos do tipo ARMA são usados quando nos modelos do tipo AR, ou do tipo MA, são necessários demasiados parâmetros para a modelação de uma determinada série temporal.

#### 3.3.1 Modelos autoregressivos

**Definição 3.3.1.** Diz-se que um processo  $\{X_t\}_{t \in \mathbb{Z}}$  admite uma representação autoregressiva de ordem  $p$ , abreviadamente  $AR(p)$ , se existirem números reais  $\phi_1, \phi_2, \dots, \phi_p$  e um processo de ruído branco  $\{\epsilon_t\}_{t \in \mathbb{Z}}$ , de média nula e de variância  $\sigma_\epsilon^2$ , e o processo  $X_t$  verificar a seguinte equação

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t \quad (\phi_p \neq 0), t \in \mathbb{Z}$$

onde  $\mu$  é uma constante. Neste caso diz-se que o processo  $X_t$  tem média não nula.

Exige-se que o coeficiente  $\phi_p$  seja diferente de zero pois é este último coeficiente que dita a ordem do processo. Por exemplo, um processo pode ser do tipo  $AR(p)$  e ser da forma  $X_t = \mu + \phi_p X_{t-p} + \epsilon_t$ . Ou seja, todos os coeficientes intermédios podem ser nulo, inclusive a constante.

Os processos  $AR(p)$  nem sempre são estacionários. Como apenas interessa o caso estacionário foram feitos estudos para avaliar em que condições estes processos são estacionários.

Para tal, defina-se o chamado *operador atraso*, designado por  $B$ , tal que

$$BX_t = X_{t-1} \quad \text{e} \quad B^k X_t = X_{t-k}$$

Então, usando o operador atraso, pode escrever-se o processo  $AR(p)$  na forma

$$\Phi(B)X_t = \mu + \epsilon_t,$$

onde  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ .

Donde se prova (Pacheco, 2001) que o processo  $AR(p)$  é estacionário se e só se o polinómio  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  não possuir raízes no círculo unitário.

Vamos agora ver algumas características deste tipo de modelos. Facilmente se prova que se o processo  $X_t$  for estacionário então

$$E[X_t] = \frac{\mu}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$$

$$Var[X_t] = \frac{\sigma_\epsilon^2}{1 - \phi_1\rho(1) - \phi_2\rho(2) - \dots - \phi_p\rho(p)}$$

E prova-se (Gomes, 2005) que a sucessão de funções de autocorrelação de um processo  $AR(p)$  estacionário verifica a seguinte condição de recorrência

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2) + \dots + \phi_p\rho(k-p), \quad k > 0,$$

Existem dois teoremas muito importantes que nos permitem através das representações gráficas das FAC e das FACP avaliar não só o tipo de modelo(s) a ajustar aos dados mas também a ordem desse(s) modelo(s).

**Teorema 3.3.1.** *A sucessão de funções de autocorrelação de um processo  $AR(p)$  estacionário tendem exponencialmente e/ou de forma sinusoidal para zero, à medida que o valor de  $k$  aumenta.*

**Teorema 3.3.2.** *A sucessão de funções de autocorrelação parcial de um processo  $AR(p)$  estacionário anulam-se a partir da ordem  $p + 1$ .*

**Exemplo 3.3.1.** *Considere-se agora o caso particular de um processo  $AR(1)$ , com a seguinte representação*

$$X_t = \mu + \phi X_{t-1} + \epsilon_t \quad (\phi \neq 0), \quad t \in \mathbb{Z}$$

onde  $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$ .

Este processo é estacionário se e só se  $|\phi| < 1$ .

Prova-se facilmente que se o processo  $X_t$  é estacionário então

$$E[X_t] = \frac{\mu}{1 - \phi}$$

$$Var[X_t] = \frac{\sigma_\epsilon^2}{1 - \phi^2}$$

$$\rho(k) = \phi^k, \quad k = 0, 1, \dots$$

donde se conclui que as FAC de um processo  $AR(1)$  estacionário tendem exponencialmente, e/ou de modo sinusoidal, para zero, à medida que o valor de  $k$  aumenta.

E que as FACP de um processo  $AR(1)$  estacionário são da forma

$$\begin{cases} \phi(1, 1) = \phi; \\ \phi(k, k) = 0, \quad k \geq 2. \end{cases}$$

Ou seja, as FACP de um processo  $AR(1)$  estacionário anulam-se a partir da ordem 2.

**Exemplo 3.3.2.** Se considerarmos agora um processo  $AR(2)$ , com a seguinte representação

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \quad (\phi_2 \neq 0), t \in \mathbb{Z}$$

$$\epsilon_t \sim RB(0, \sigma_\epsilon^2).$$

Prova-se (Gomes, 2005) que um processo  $AR(2)$  é estacionário se e só se

$$\phi_2 + \phi_1 < 1, \quad \phi_2 - \phi_1 < 1 \quad \text{e} \quad |\phi_2| < 1$$

Se o processo  $X_t$  for estacionário após alguns cálculos obtém-se que

$$E[X_t] = \frac{\mu}{1 - \phi_1 - \phi_2}$$

$$Var[X_t] = \frac{\sigma_\epsilon^2}{1 - \phi_1 \rho(1) - \phi_2 \rho(2)}$$

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2), \quad k > 0$$

E obtém-se os seguintes teoremas (Gomes, 2005)

**Teorema 3.3.3.** A sucessão de funções de autocorrelação de um processo  $AR(2)$  estacionário tendem exponencialmente, e/ou de modo sinusoidal, para zero, à medida que o valor de  $k$  aumenta.

E também que

$$\begin{cases} \phi(1, 1) = \frac{\phi_1}{1 - \phi_2} \\ \phi(2, 2) = \frac{\rho(2) - \rho^2(1)}{1 - \rho(1)^2} \\ \phi(k, k) = 0, \quad k \geq 3 \end{cases}$$

Ou seja,

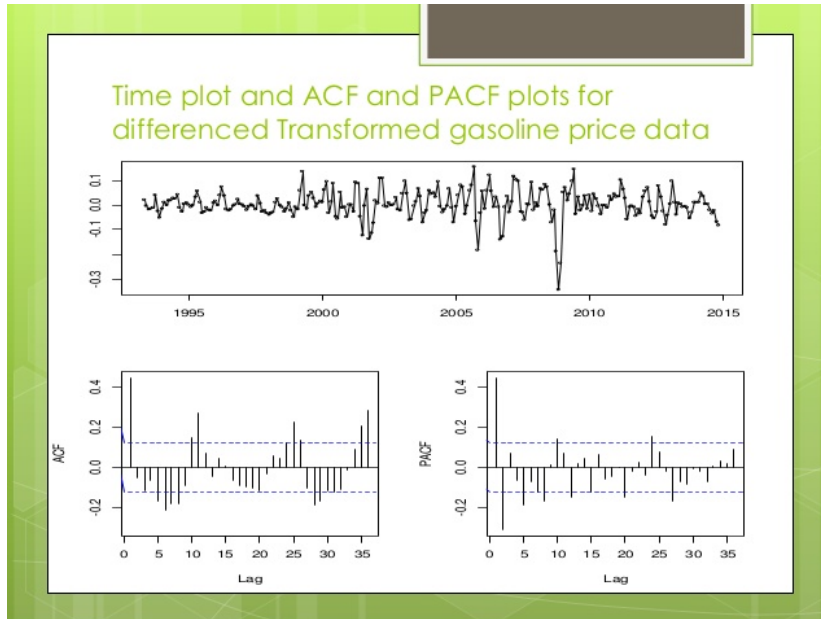
**Teorema 3.3.4.** A sucessão de funções de autocorrelação parcial de um processo  $AR(2)$  estacionário anulam-se a partir da ordem 3.

No exemplo seguinte apresentamos uma série temporal (transformada) do preço da gasolina. Iremos na secção 3.4 introduzir o caso das séries temporais lineares não estacionárias e das transformações necessárias para as transformar em estacionárias. Apresenta-se também nesta figura as funções de autocorrelação (FAC) e autocorrelação parcial (FACP). Como se pode observar, a FAC tem um decaimento sinusoidal, para zero, à medida que o valor de  $k$  aumenta e a FACP anulam-se ou a partir da ordem 3 ou a partir da ordem 5. Pelo que temos dois modelos possíveis para modelar esta série temporal, um modelo  $AR(2)$  e um modelo  $AR(5)$ . Mais à frente iremos averiguar como avaliar (testar) qual dos dois modelos é o mais adequado para modelar a série e também testar se a série tem média nula ou não. Ou seja, se o modelo inclui a constante  $\mu$  ou não.

### 3.3.2 Modelos de médias móveis

**Definição 3.3.2.** Diz-se que um processo  $\{X_t\}_{t \in \mathbb{Z}}$  admite uma representação média móvel de ordem  $q$ , abreviadamente  $MA(q)$ , se existirem números reais  $\theta_1, \theta_2, \dots, \theta_q$  e um processo de ruído branco  $\{\epsilon_t\}_{t \in \mathbb{Z}}$ , de variância  $\sigma_\epsilon^2 > 0$ , tais que o processo  $X_t$  verifica a seguinte equação

$$X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (\theta_q \neq 0), t \in \mathbb{Z}$$



**Figura 3.4:** Série temporal, FAC e FACP de um AR(2)

(<https://www.slideshare.net/SrinivasKNS1/presentation-42950288>)

com  $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$  ( $\sigma_\epsilon^2 > 0$ ).

**Nota:** Um processo MA( $q$ ) é sempre um processo estacionário, quaisquer que sejam os valores das constantes  $\theta_1, \dots, \theta_q$ , uma vez que é definido apenas à custa de um ruído branco.

Então facilmente se prova que

$$E[X_t] = \mu,$$

$$Var[X_t] = \sigma_\epsilon^2 \left(1 + \sum_{i=1}^q \theta_i^2\right),$$

e que (Gomes, 2005)

$$\rho(k) = \frac{\sum_{i=1}^{q-k} \theta_i \theta_{i+k} - \theta_k}{1 + \sum_{i=1}^q \theta_i^2}, \quad k = 0, 1, \dots, q \quad \rho(k) = 0, \quad k > q$$

**Teorema 3.3.5.** A sucessão de funções de autocorrelação de um processo MA( $q$ ) estacionário anulam-se a partir da ordem  $q + 1$ .

**Teorema 3.3.6.** A sucessão de funções de autocorrelação parcial de um processo MA( $q$ ) estacionário tendem exponencialmente e/ou de forma sinusoidal para zero, à medida que o valor de  $k$  aumenta.

**Exemplo 3.3.3.** Seja  $X_t$  um processo MA(2) com a seguinte representação

$$X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} \quad (\theta_2 \neq 0), \quad t \in \mathbb{Z}$$

com  $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$ .

Então

$$E[X_t] = \mu$$

$$Var[X_t] = \sigma_\epsilon^2(1 + \theta_1^2 + \theta_2^2),$$

e

$$\begin{cases} \rho(1) = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \\ \rho(2) = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \\ \rho(k) = 0, k \geq 3 \end{cases}$$

No exemplo seguinte apresentamos uma série temporal à qual se ajusta bem um modelo do tipo MA(2). Ou seja, a FAC anula-se a partir da ordem 3 (isto é, apenas os dois primeiros valores estão fora dos limites de confiança) e a FACP tendem exponencialmente para zero à medida que o valor de  $k$  aumenta.

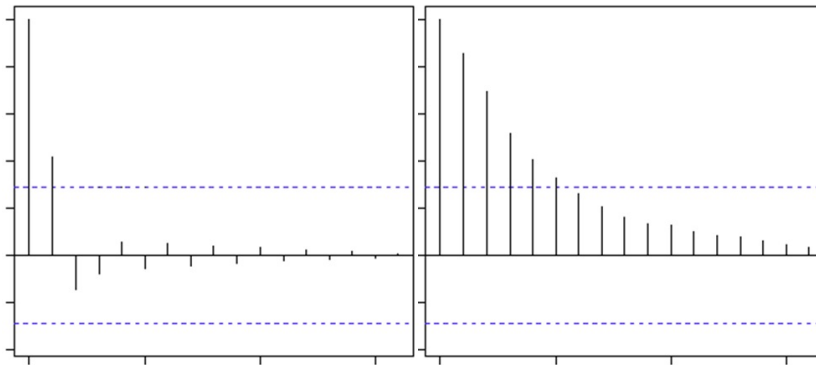


Figura 3.5: Série temporal, FAC e FACP de um MA(2) (dados simulados)

### 3.3.3 Modelos autoregressivos e de médias móveis

Os modelos autoregressivos e de médias móveis são construídos através da combinação dos dois modelos anteriores, AR e MA. Tal como foi referido anteriormente, os modelos do tipo ARMA são usados muitas vezes quando nos modelos do tipo AR, ou do tipo MA, são necessários demasiados parâmetros para a modelação de uma série temporal.

**Definição 3.3.3.** Diz-se que um processo  $\{X_t\}_{t \in \mathbb{Z}}$  admite uma representação autoregressiva média móvel de ordem  $(p, q)$ , abreviadamente ARMA( $p, q$ ), se existirem números reais  $(\phi_1, \phi_2, \dots, \phi_p)$ ,  $(\theta_1, \theta_2, \dots, \theta_q)$  e um processo de ruído branco  $\{\epsilon_t\}_{t \in \mathbb{Z}}$ , de variância  $\sigma_\epsilon^2 > 0$ , tais que o processo  $X_t$  verifica a seguinte equação

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (\phi_p, \theta_q \neq 0) \quad (3.1)$$

com  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ ,  $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  e  $\phi_1, \dots, \phi_p$  e  $\theta_1, \dots, \theta_q$  sem raízes comuns.

Se  $X_t$  admitir uma representação do tipo ARMA( $p, q$ ) ele é estacionário se e só se o polinómio  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  não possuir raízes no círculo unitário.

Após alguns cálculos prova-se que



Então

$$E[X_t] = \frac{\mu}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$$

Prova-se (Pacheco, 2001) que a função de autocorrelação de um modelo ARMA( $p, q$ ) verifica a seguinte equação

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2) + \dots + \phi_p\rho(k-p), \quad k \geq q+1.$$

**Teorema 3.3.7.** *A sucessão de funções de autocorrelação de um processo ARMA( $p, q$ ) estacionário tendem exponencialmente e/ou de forma sinusoidal para zero, à medida que o valor de  $k$  aumenta.*

**Teorema 3.3.8.** *A sucessão de funções de autocorrelação parcial de um processo ARMA( $p, q$ ) estacionário tendem exponencialmente e/ou de forma sinusoidal para zero, à medida que o valor de  $k$  aumenta.*

**Exemplo 3.3.4.** *Seja  $X_t$  um processo ARMA(1, 1) com a seguinte representação*

$$X_t = \mu + \phi X_{t-1} + \epsilon_t - \theta \epsilon_{t-1}, \quad t \in \mathbb{Z}$$

com  $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$ .

Neste caso,  $X_t$  é estacionário se e só se  $|\phi| < 1$ .

Prova-se facilmente que se o processo  $X_t$  é estacionário então

$$E[X_t] = \frac{\mu}{1 - \phi},$$

$$Var[X_t] = \frac{\sigma_\epsilon^2(1 - 2\phi\theta + \theta^2)}{1 - \phi^2}$$

e

$$\begin{cases} \gamma(1) = \phi\gamma(0) - \theta\sigma_\epsilon^2 \\ \gamma(k) = \phi\gamma(k-1), \quad k \geq 2 \end{cases}$$

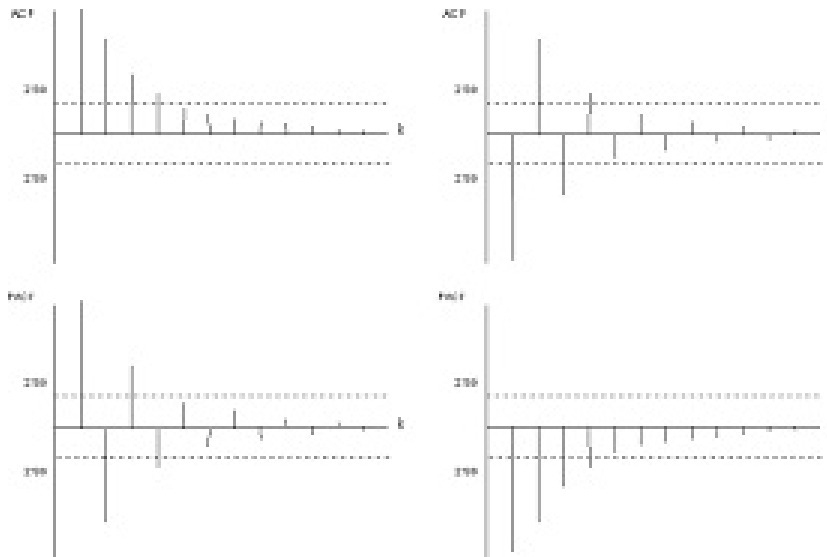
Tal que  $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ ,  $k \geq 1$ .

No exemplo seguinte apresentamos uma série temporal que pode ser bem ajustada por um modelo do tipo ARMA(1, 1). Onde se observa que tanto a FAC como a FACP tendem exponencialmente para zero à medida que o valor de  $k$  aumenta.

### 3.4 Modelos para séries temporais lineares não estacionárias

Em muitos casos, é usual encontrar processos que são não-estacionários, e esses acontecem quando existem tendências ou/e uma sazonalidade dos dados, que normalmente são causados pela não constância quer da média quer da variância/covariância.

Como forma de transformar um processo não-estacionário em estacionário, existem técnicas que possibilitam essa transformação, como também modelos para processos não-estacionários. Quando



**Figura 3.6:** Série temporal, FAC e FACP de um ARMA(1,1) ([https://www.xycoon.com/arma11\\_process.htm](https://www.xycoon.com/arma11_process.htm))

uma série temporal só apresenta tendência elimina-se esta através de técnicas para estabilizar a variância (*chamadas transformações de Box-Cox*) e depois fazem-se diferenciações simples, tantas quantas as necessárias.

Quando uma série temporal apresenta só sazonalidade pode eliminar-se esta e transformar a série não estacionária numa série estacionária através de diferenciações sazonais, também tantas quantas as necessárias.

Quando uma série apresenta tendência e sazonalidade deve eliminar-se sempre primeiro a tendência e só depois a sazonalidade (Gomes, 2015).

### 3.4.1 Transformações para estabilizar a variância

A metodologia mais usual para estabilizar a variância de uma série temporal são as chamadas *transformações de Box-Cox* (de parâmetro  $\lambda$ ). Se  $\{X_t\}_{t \in \mathbb{Z}}$  for uma série não estacionária, então a transformação de Box-Cox consiste numa transformação,  $T$ , aplicada à série do seguinte modo:

$$T(X_t)_\lambda = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & X_t \geq 0, \lambda \neq 0 \\ \ln(X_t), & X_t > 0, \lambda = 0 \end{cases}$$

Normalmente, o valor de  $\lambda \in [-1, 1[$ .

Quando se usa uma transformação de Box-Cox todos os dados devem ser positivos. Contudo, mesmo as séries tenha valores negativos é possível aplicar essas transformações pois basta adicionar uma dada constante positiva aos dados para que estes se tornem positivos.

Admitamos que estamos perante uma dada realização, de dimensão  $n$ , do p.e.  $\{X_t\}_{t \in \mathbb{Z}}$  de valores positivos, onde a variância do processo não sendo constante, é função do respectivo valor médio. Neste caso, teríamos grandes dificuldades em saber o valor de  $\lambda$  mais adequado para a transformação (Gomes, 2005).

Uma forma de se saber o valor de  $\lambda$  mais adequado para a transformação, seria supor que  $X_t, t = 1, 2, \dots, n$  uma realização, de dimensão  $n$ , do processo  $X_t$ , e realizar o cálculo do seguinte coeficiente

$$R_\lambda(X_t) = \begin{cases} X_t^\lambda \bar{X}_t / \bar{X}_t^\lambda, & X_t \geq 0, \lambda \neq 0 \\ \ln(X_t) \bar{X}_t / \ln(\bar{X}_t), & X_t > 0, \lambda = 0 \end{cases}$$

com  $X_t^\lambda$  a série transformada e  $\bar{X}_t^\lambda$  a sua média amostral. O valor de  $\lambda$  mais adequado é aquele que conduz a uma menor variância amostral para a série transformada. Neste caso, aquele que acarreta

menor valor de  $S_{R,\lambda}^2(X_t)$ . Depois da escolha do valor de  $\lambda$ , podemos cita-lo como  $\lambda^*$ , pode-se utilizar a seguinte transformação para estabilizar a variância

$$T(X_t)_\lambda^* = \begin{cases} X_t^{\lambda^*}, & X_t \geq 0, \lambda \neq 0 \\ \ln(X_t), & X_t > 0, \lambda = 0. \end{cases}$$

Depois da variância estabilizada, passa-se para o próximo passo que é a estabilização da média.

### 3.4.2 Transformações para eliminar as tendências

Normalmente, depois da estabilização da variância, pode suceder que os dados apresentem uma forte tendência de crescimento ou decrescimento, que pode ser visível através do declínio lento da FAC empírica, o que indica, que a média do processo não é constante (Gomes, 2005).

Se consideramos que  $\{Y_t\}_{t \in \mathbb{Z}}$  é a série resultante do processo de estabilização da variância (ou seja,  $Y_t = T(X_t)$ ), é possível eliminar a tendência de uma dada série temporal através da aplicação do *operador diferença simples* de ordem  $d$  (onde  $d$  representa o número de diferenças simples necessárias para estabilizar a tendência). O caso mais simples, e muitas vezes o mais usual, é quando  $d = 1$  e aplica-se do seguinte modo:

$$DY_t = Y_t - Y_{t-1}, \quad t \in \mathbb{Z}$$

Usualmente também se costuma escrever as diferenciações à custa do operador atraso  $B$ . Ou seja,  $DY_t = (1 - B)Y_t$ , onde  $BY_t = Y_{t-1}$ .

Por vezes é preciso aplicar o operador diferença mais de uma vez para poder eliminar a tendência de uma série. Mas deverá ser aplicado somente o número de vezes necessário para estabilizar a média, caso contrário, poderá estar introduzir-se uma variância adicional à série, estando igualmente a eliminar informação (Gomes, 2005).

O operador diferença com ordens maiores ( $d > 1$ ) é definido passo a passo. Por exemplo, para  $d = 2$ , vem

$$D^2Y_t = D(DY_t) = D(Y_t - Y_{t-1}) = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2}), \quad t \in \mathbb{Z}$$

ou seja,

$$D^2Y_t = Y_t - 2Y_{t-1} + Y_{t-2}, \quad t \in \mathbb{Z}$$

Outra forma usual de o representar é do seguinte modo:  $(1 - B)^2Y_t$ .

Para estabilizar a componente sazonal aplica-se o *operador diferença sazonal* de ordem  $D$  (onde  $D$  representa o número de diferenças sazonais necessárias para estabilizar a componente sazonal) com período de sazonalidade  $S$ . Considerando o caso também mais simples e mais usual, quando  $D = 1$ , e para um período de sazonalidade  $S$ , este operador define-se do seguinte modo:

$$D_S Y_t = Y_t - Y_{t-s}, \quad t \in \mathbb{Z}$$

Ou  $D_S Y_t = (1 - B^S)Y_t$ . Para ordens  $D > 1$  procede-se da mesma maneira que anteriormente. Ou seja,  $D_S^D Y_t = (1 - B^S)^D Y_t$ .

Quando uma série temporal apresentar tendência e sazonalidade e for necessário aplicar os operadores diferença, deve aplicar-se primeiro o operador diferença sazonal e só depois o operador diferença simples (Gomes, 2005).

### 3.4.3 Modelo ARIMA

Os modelos integrados autoregressivos e médias móveis de ordem  $(p,d,q)$  ( $ARIMA(p, d, q)$ ) são modelos lineares propostos para modelar séries temporais não estacionárias com tendência, onde AR:  $p$ =grau da parte autoregressiva, I:  $d$ =ordem do operador diferença simples e MA:  $q$ =grau da parte de média móvel (Renato Sato, 2013). O modelo  $ARIMA(p, d, q)$  não-estacionário pode ser transformado num modelo  $ARMA(p, q)$ , caso ele seja diferenciado  $d$  vezes através do operador diferença simples (Gomes, 2005).

**Definição 3.4.1.** Diz-se que um processo  $\{X_t\}_{t \in \mathbb{Z}}$  admite uma representação integrada autoregressiva e médias móveis de ordem  $(p,d,q)$ , abreviadamente  $ARIMA(p,d,q)$ , se existirem números reais  $(\phi_1, \phi_2, \dots, \phi_p)$ ,  $(\theta_1, \theta_2, \dots, \theta_q)$  e um processo de ruído branco  $\{\epsilon_t\}_{t \in \mathbb{Z}}$ , de variância  $\sigma_\epsilon^2 > 0$ , tais que o processo  $X_t$  verifica a seguinte equação

$$\Phi(B)(1-B)^dT(X_t) = \mu + \Theta(B)\epsilon_t,$$

onde  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  e  $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  são, respectivamente, o polinómio autoregressivo e o polinómio médias móveis e  $T$  é a transformação para estabilizar, se necessário, a variância. Os polinómios  $\Phi(B)$  e  $\Theta(B)$  não têm raízes no círculo unitário e não têm raízes comuns (Gomes, 2005).

O processo  $ARIMA(p,d,q)$  não é mais do que uma generalização dos processos  $ARMA(p,q)$  quando  $d = 0$ .

Quando  $p = 0$  processo  $\{X_t\}_{t \in \mathbb{Z}}$  é representado da seguinte forma

$$(1-B)^dT(X_t) = \mu + \Theta(B)\epsilon_t,$$

designado processo integrado de médias móveis de ordem  $(d,p)$ , abreviadamente  $IMA(d,q)$ .

De maneira similar, quando  $q = 0$  o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é representado da seguinte forma

$$\Phi(B)(1-B)^dT(X_t) = \mu + \epsilon_t$$

designado processo integrado autoregressivo de ordem  $(p,d)$ , abreviadamente  $ARI(p,d)$ .

### 3.4.4 Modelo SARIMA

Normalmente as séries económicas e as séries associadas aos processos físicos e ambientais apresentam componente sazonal. O modelo  $SARIMA(p,d,q) \times (P,D,Q)_S$  foi desenvolvido pensando no conjunto de dados que pode possuir tendência e/ou uma componente sazonal de período  $S$ . Ou seja, trata-se de uma generalização do modelo  $ARMA(p,q)$ .

**Definição 3.4.2.** Diz-se que um processo  $\{X_t\}_{t \in \mathbb{Z}}$  admite uma representação mista integrada autoregressiva e médias móveis e de componente sazonal de período  $S$  de ordens  $(p,d,q)$  e  $(P,D,Q)$ , abreviadamente  $SARIMA(p,d,q) \times (P,D,Q)_S$ , se existirem números reais  $(\phi_1, \phi_2, \dots, \phi_p)$ ,  $(\theta_1, \theta_2, \dots, \theta_q)$ ,  $(\Phi_1, \Phi_2, \dots, \Phi_P)$ ,  $(\Theta_1, \Theta_2, \dots, \Theta_Q)$  e um processo de ruído branco  $\{\epsilon_t\}_{t \in \mathbb{Z}}$ , de variância  $\sigma_\epsilon^2 > 0$ , tais que o processo  $X_t$  verifica a seguinte equação

$$\Delta(B^S)\Phi(B)(1-B)^dT(X_t) = \mu + \Psi(B^S)\Theta(B)\epsilon_t, \quad t \in \mathbb{Z}$$

onde  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ ,  $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  são, respectivamente, os polinómios autoregressivo e de médias móveis da parte não sazonal e  $\Delta(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_P B^P$  e  $\Psi(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_Q B^Q$  são, respectivamente, os polinómios autoregressivo e de médias móveis da parte sazonal.  $T$  é a transformação para estabilizar, se necessário, a variância. De igual modo, todos os polinómios não têm raízes no círculo unitário e não têm raízes comuns (Gomes, 2005).

A parte não sazonal do modelo  $SARIMA(p,d,q) \times (P,D,Q)_S$  é representada pelos parâmetros  $(p,d,q)$  e a parte sazonal pelos parâmetros  $(P,D,Q)$ . O objectivo é estabelecer uma relação entre estas duas partes, sendo que a primeira estabelece uma relação dentro dos períodos e a segunda entre os períodos (Gomes, 2005). Segundo (Murteira, 1993) a ordem das diferenciações normalmente é  $d \geq 0$  e  $D \leq 2$  e os valores de  $p, q, P$  e  $Q$  raramente ultrapassam o valor 2.

A aplicação dos operadores diferenças simples e/ou sazonais é uma ferramenta muito importante e eficaz para tornar uma série não estacionária numa série estacionária, mas a sua desvantagem

é que acarreta custos na altura da previsão, isto porque sempre que se fazem diferenciações perdem-se dados. Por exemplo, se tivermos uma série com dados mensais ao longo de alguns anos e esta tiver uma componente sazonal, de período anual ( $S = 12$ ), e uma componente de tendência e se fizermos  $d = D = 1$  iremos perder 13 observações (um ano inteiro mais um mês).

Neste sentido, deve equacionar-se usar modelos ARMA estritamente sazonais de período  $S$  e com ordens  $P$  e  $Q$ , que conseguem captar a sazonalidade e que não precisam de diferenciações. Estes modelos são da forma

$$\Delta(B^S)T(X_t) = \mu + \Psi(B^S)\epsilon_t, \quad t \in \mathbb{Z}$$

com  $\Delta(B)$  e  $\Psi(B)$  os polinómios atrás referidos e  $\{\epsilon_t\}_{t \in \mathbb{Z}}$ , um de ruído branco de variância  $\sigma_\epsilon^2 > 0$  (Pacheco, 2001).

## 3.5 Modelação de séries temporais

Nesta secção iremos abordar a modelação de série temporais lineares, estacionárias ou não estacionárias.

A modelação de uma série temporal, ou seja, a identificação de um (ou mais) modelo(s) para a sua caracterização implica os seguintes passos:

1. Série estacionária → Identificação do modelo → Estimação dos parâmetros do modelo → Avaliação do diagnóstico
2. Série não estacionária → Transformações → Identificação do modelo → Estimação dos parâmetros do modelo → Avaliação do diagnóstico

### 3.5.1 Identificação do modelo

A primeira etapa antes da identificação do modelo consiste em saber se a série  $\{X_t \mid t = 1, 2, \dots, n\}$  é ou não estacionária.

Existem alguns testes estatísticos para averiguar se a série é ou não estacionária mas uma maneira muito usual é através das representações gráficas do cronograma e da FAC. Por vezes a não estacionariedade é muito visível através do cronograma. Outras vezes pode não ser assim tão evidente. Mas, em geral, quando a FAC tende muito lentamente para zero isso significa que a série não é estacionária (Gomes, 2005).

No caso da série não ser estacionária aplicam-se as transformações descritas na secção anterior. Ou seja, começa-se por estabilizar a variância através da transformação de Box-Cox e depois aplicam-se os operadores diferença simples e/ou sazonal. Sendo que se for necessário aplicarem-se as duas diferenciações primeiro aplica-se a diferenciação sazonal e só depois a simples.

Caso a série seja estacionária, ou depois da série ter ficado estacionária através das transformações aplicadas, iremos então identificar o(s) modelo(s) mais adequados. Depois iremos identificar as suas ordens  $p$ ,  $q$ ,  $P$  e  $Q$ . Tal como vimos na secção 3.3, as ordens  $p$  e  $q$  dos modelos ARMA são identificadas através do comportamento conjunto das FAC e FACP. Do mesmo modo as ordens  $P$  e  $Q$  também são identificadas do mesmo modo, mas entre os períodos. Ou seja, se  $P \geq 0$  e/ou  $Q \geq 0$  e a série tiver período  $S$ , os valores de  $p$  e  $q$  são identificados, respectivamente, através das ordens (ou lags)  $1, 2, \dots, S - 1$  das FAC e FACP. Enquanto que os valores de  $P$  e  $Q$  são identificados, respectivamente, através das ordens  $S, 2S, 3S, \dots$ , múltiplas do período.

Para determinar qual o tipo de modelo, ou seja se estamos perante um modelo do tipo AR, MA ou ARMA, é necessário ter em consideração os Teoremas 3.3.1-3.3.6 apresentados na secção 3.3. Depois de identificado o(s) tipo(s) de modelo(s) é então preciso determinar as suas ordens.

Para determinar quais os valores da FAC que são nulos faz-se o seguinte teste de hipóteses, admitindo que a série é uma realização de um processo Gaussiano,

$$H_0 : \rho(k) = 0 \quad vs \quad H_1 : \rho(k) \neq 0$$

para  $k = 1, 2, \dots$  e  $\rho(k)$  a FAC do processo.

Se o processo for estacionário e Gaussiano prova-se que a variância do estimador da FAC é aproximadamente  $1/n$ . Ou seja,

$$\text{Var}[\hat{\rho}(k)] \simeq \frac{1}{n}$$

Então, para testar a hipótese  $H_0$  usa-se a seguinte estatística de teste

$$Z = \frac{\hat{\rho}(k)}{\sqrt{\text{Var}[\hat{\rho}(k)]}} \simeq \sqrt{n}\hat{\rho}(k) \simeq N(0, 1)$$

Então não se rejeita-se  $H_0$ , ao nível de significância  $\alpha$  se  $|\sqrt{n}\hat{\rho}(k)| < Z_{1-\alpha/2}$ , onde  $Z_{1-\alpha/2}$  é o quantil de probabilidade  $1 - \alpha/2$  da distribuição normal padrão (Gomes, 2005).

Da mesma forma, para identificar quais os valores da FACP que são nulos deverá proceder-se ao seguinte teste de hipótese

$$H_0 : \phi(k, k) = 0 \quad \text{vs} \quad H_1 : \phi(k, k) \neq 0$$

com a seguinte estatística de teste

$$Z = \sqrt{n}\phi(k, k) \simeq N(0, 1)$$

A não rejeição de  $H_0$  faz-se do mesmo modo que no teste anterior.

Tal como referimos na subsecção 3.2.3., os gráficos das FAC e FACP apresentam linhas horizontais a tracejado que representam os intervalos de confiança (em geral a 95%) destas funções. Como os testes de hipóteses anteriores são bilaterais existem uma correspondência directa entre estes e os intervalos de confiança. Ou seja, os valores em cada  $k$ , quer da FAC quer da FACP, que estiverem dentro dos intervalos são considerados nulos, o que equivale a não rejeitar  $H_0$  para esse nível de confiança.

### 3.5.2 Estimação dos parâmetros do modelo

Depois de identificado o(s) modelo(s), que à simples vista parecem ser os que vão descrever melhor a série, passa-se à fase seguinte, que é a da estimação dos parâmetros do(s) modelo(s). Como ficaria muito extenso abordar todos os estimadores para todos os modelos anteriormente vistos neste capítulo, optou-se por abordar os estimadores usuais apenas para um modelo AR( $p$ ). Em caso de necessidade pode consultar-se outros estimadores em, por exemplo, (Gomes, 2005) e (Ehlers, 2007).

Seja então  $\{X_t\}_{t \in \mathbb{Z}}$  um modelo AR( $p$ ) estacionário e de média não nula dado por

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t, \quad t \in \mathbb{Z}$$

com  $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$ .

#### 3.5.2.A Estimador dos momentos para o modelo AR( $p$ )

O estimador dos momentos dos parâmetros do modelo AR( $p$ ) é obtido através da equações de Yule-Walker, substituindo  $\rho_k$  (autocorrelações teóricas) por  $\hat{\rho}_k$  (autocorrelações estimadas) no sistema de equações (Gomes, 2005)

$$\begin{cases} \rho(1) - \phi_1 - \phi_2 \rho(1) - \dots - \phi_p \rho(1-p) = 0 \\ \rho(2) - \phi_1 \rho(1) - \phi_2 - \dots - \phi_p \rho(2-p) = 0 \\ \rho(p) - \phi_1 \rho(p-1) - \phi_2 \rho(p-2) - \dots - \phi_p = 0 \end{cases}$$

Por este motivo estes estimadores dos momentos também são conhecidos por estimadores de Yule-Walker.

Depois de obtidos os estimadores do vector de parâmetros  $(\phi_1, \dots, \phi_p)$  o estimador da constante,  $\mu$ , e da variância dos erros,  $\sigma_\epsilon^2$ , podem ser obtidos do seguinte modo

$$\hat{\mu} = (1 - \hat{\phi}_1 - \hat{\phi}_2 - \dots - \hat{\phi}_p) \bar{X}_t$$

e

$$\hat{\sigma}_\epsilon^2 = S^2(1 - \hat{\phi}_1 \hat{\rho}(1) - \hat{\phi}_2 \hat{\rho}(2) - \dots - \hat{\phi}_p \hat{\rho}(p))$$

### 3.5.2.B Estimador dos mínimos quadrados condicionais para o modelo AR(p)

O estimador dos mínimos quadrados condicionais (admite-se que os valores iniciais da série temporal e/ou da série de ruído branco associada são conhecidos) dos parâmetros  $\vec{\vartheta} = (\phi_1, \phi_2, \dots, \phi_p, \mu)$  do modelo AR(p), dada uma série temporal de dimensão  $n$ ,  $(X_1, \dots, X_n)$ , consiste em minimizar a soma dos quadrados

$$Q_n(\vec{\vartheta}) = \sum_{t=p+1}^n (X_t - \mu - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p})^2$$

Particularizando ao modelo AR(1)S prova-se (Gomes, 2005) que desta minimização resultam, respectivamente, os seguintes estimadores dos mínimos quadrados para  $\phi$  e  $\mu$ ,

$$\hat{\phi} = \frac{\sum_{t=2}^n (X_t - \bar{X})(X_{t-1} - \bar{X})}{\sum_{t=2}^n (X_{t-1} - \bar{X})^2}$$

e

$$\hat{\mu} = \frac{1}{n-1} \sum_{t=2}^n X_t - \hat{\phi} \left\{ \frac{1}{n-1} \sum_{t=2}^n X_{t-1} \right\}$$

Pode ainda estimar-se  $\sigma_\epsilon^2$  a partir da estimação de  $\phi$  e através do estimador usual da variância,

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-2} \sum_{t=2}^n (X_t - \hat{\mu} - \hat{\phi} X_{t-1})^2$$

**Nota:** Para uma dimensão da amostra suficientemente grande, os estimadores dos mínimos quadrados condicionais e os estimadores de Yule-Walker de um processo AR(p) pouco ou nada diferem uns dos outros. Caso algum dos parâmetros do processo esteja próximo de um, a estimação de Yule-Walker deverá ser evitada uma vez que fornece más estimativas para os parâmetros (Gomes, 2005).

### 3.5.2.C Estimador de máxima verosimilhança condicionada para o modelo AR(p)

A função de verosimilhança do processo AR(p)S,  $L(\vec{\vartheta}) = L(\vec{x}_n; \vec{\vartheta})$ , onde  $\vec{\vartheta}$  é o vector de parâmetros desconhecidos  $(\phi_1, \phi_2, \dots, \phi_p, \mu)$ , escreve-se do seguinte modo

$$L(\vec{\vartheta}) = f(x_1) f(x_2|x_1) \dots f(x_{n-1}|x_{n-2}, \dots, x_1) f(x_n|x_{n-1}, \dots, x_1).$$

Como em geral não se conhece a distribuição de  $X_t$  pode considerar-se  $X_1$  é conhecido. Deste modo, proceder-se à maximização da *verosimilhança condicionada* à primeira observação, ou seja,

$$L(\vec{\vartheta}) = f(x_2|x_1) \dots f(x_{n-1}|x_{n-2}, \dots, x_1) f(x_n|x_{n-1}, \dots, x_1)$$

Considerando que  $\epsilon_t$  têm distribuição normal, então estas v.a.'s são independentes e têm a mesma distribuição.

Prova-se (Pacheco, 2001), então que a função de verosimilhança, condicionada aos valores observados  $(x_1, x_2, \dots, x_p)$ , do processo escreve-se do seguinte modo

$$L(\vec{\vartheta}) = \left( \frac{1}{2\pi\sigma_\epsilon^2} \right)^{\frac{n-p}{2}} \exp \left[ -\frac{1}{2\sigma_\epsilon^2} \sum_{t=p+1}^n \epsilon_t^2 \right]$$

E prova-se ainda, após alguns cálculos, que

$$\log(L(\vec{\vartheta})) = -(n-p) \log(2\pi\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \sum_{t=p+1}^n (X_t - \mu - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p})^2$$

donde se retira facilmente que este logaritmo é máximo quando a soma dos quadrados

$$\sum_{t=p+1}^n (X_t - \mu - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p})^2$$

for mínima.

**Nota:** Assim, obter os estimadores de máxima verosimilhança condicional de um processo  $AR(p)$ , considerando que as inovações (ou os erros)  $\epsilon_t$  têm distribuição normal, corresponde a obter os estimadores dos mínimos quadrados condicionais. (Gomes, 2005).

### 3.5.3 Avaliação do diagnóstico

Após identificados o(s) modelo(s), as respectivas ordens e estimados os seus parâmetros, deve realizar-se um diagnóstico para verificar a sua adequação, antes de utilizá-lo para fazer previsões. A avaliação da adequação do(s) modelo(s) faz-se através da avaliação das seguintes características (Pacheco, 2001):

1. Significância estatística dos parâmetros do modelo;
2. Estacionariedade e invertibilidade do modelo;
3. Estabilidade do modelo;
4. Redundância dos parâmetros do modelo;
5. Avaliação da qualidade do modelo.

#### 3.5.3.A Significância estatística dos parâmetros do modelo

A avaliação da significância estatística dos parâmetros do modelo consiste em fazer testes de hipóteses aos parâmetros do modelo, que permitem rejeitar ou não, através de uma regra decisória, uma hipótese estatística com base nos resultados da amostra.

Estas hipóteses são sobre os parâmetros populacionais e a realização dos testes é baseada na distribuição amostral dos respectivos estimadores.

Considerando, por exemplo, um modelo  $ARMA(p, q)$  com os seguintes parâmetros estimados,  $\hat{\beta}_i = (\hat{\mu}, \hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)$ , pretende-se saber qual(ais) deste(s) parâmetro(s) deve(m) ser eliminado(s) do modelo.

Ou seja, consiste em testar as seguintes hipóteses

$$H_0 : \beta_i = 0 \quad \text{vs} \quad H_1 : \beta_i \neq 0, \quad i = 1, 2, \dots, p + q + 1$$

através da seguinte estatística de teste

$$T = \sqrt{n} \frac{\hat{\beta}_i}{\sqrt{\hat{Var}[\hat{\beta}_i]}} \sim t_{(n-(p+q+1))}$$

Deve rejeitar-se a hipótese  $H_0$ , para um nível de significância  $\alpha$ , sempre que  $|T| > t_{(n-(p+q+1)), \alpha/2}$ , com  $t_{(n-(p+q+1)), \alpha/2}$  o quantil de probabilidade  $1-\alpha/2$  da distribuição t-student com  $(n-p-q-1)$  graus de liberdade. A não rejeição da hipótese nula, para um nível de significância  $\alpha$ , deverá implicar à eliminação do parâmetro  $\beta_i$  do modelo (Pacheco, 2001).

#### 3.5.3.B Estacionariedade e invertibilidade do modelo

Por exemplo, se considerarmos um modelo  $AR(1)$  vimos que este é estacionário se  $|\phi| < 1$ . Então, para avaliar se o processo é estacionário deverá testar-se a seguinte hipótese

$$H_0 : \phi = 1 \quad \text{vs} \quad H_1 : \phi < 1$$

Para mais informação consultar, por exemplo, (Pacheco, 2001).



### 3.5.3.C Estabilidade do modelo

Para que o modelo selecionado seja adequado para descrever a série em estudo o mesmo tem que passar por todos os teste de diagnóstico. Após ter passado as etapas anteriores, a próxima será saber se o mesmo é igualmente adequado para descrever o comportamento futuro desta série. Daí surgir a questão da estabilidade do modelo.

Considera-se um modelo estável quando as estimativas dos parâmetros do modelo escolhido, ao serem estimadas como base numa outra realização do processo, pouco diferem das anteriormente obtidas. *Ou seja, existe correlação entre os estimadores dos parâmetros mas essa correlação é pouco relevante. Em geral, considera-se que os estimadores se encontram fracamente correlacionados quando o coeficiente de correlação for, em módulo, inferior a 0,7.* (Gomes, 2005). Se o modelo se revelar instável deve-se procurar outros modelos alternativos aos propostos.

### 3.5.3.D Redundância dos parâmetros do modelo

Segundo o princípio da parcimónia, o modelo mais adequado deverá ser aquele que, passado na avaliação do diagnóstico, tiver o menor número possível de parâmetros (Gomes, 2005).

Assim, se pensarmos num modelo ARMA( $p, q$ ) da forma

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (\phi_p, \theta_q \neq 0)$$

ou escrito na forma

$$\Phi(B)X_t = \Theta(B)\epsilon_t$$

E agora num outro modelo ARMA( $p+1, q+1$ ), com mais parâmetros que o anterior, mas que se consegue escrever na forma

$$(1 - \alpha_1 B)\Phi(B)X_t = (1 - \alpha_2 B)\Theta(B)\epsilon_t$$

Se  $\alpha_1 \simeq \alpha_2$  então temos uma redundância entre os dois modelos (com  $1 - \alpha B$  o factor de redundância). *Esta redundância de parâmetros em modelos ARMA( $p, q$ ) tem a ver com o facto de ambos os modelos serem praticamente equivalentes (possuem FAC e FACP muito parecidas). E, neste caso, segundo o princípio da parcimónia, escolhia-se o modelo ARMA( $p, q$ ) porque era o modelo com menos parâmetros* (Gomes, 2005).

### 3.5.3.E Avaliação da qualidade do modelo

Após as análises anteriores à qualidade do modelo ajustado deve agora averiguar-se se ele cumpre com os pressupostos teóricos do mesmo. Essa averiguação faz-se através do estudos dos resíduos estimados.

Considerando um modelo ARMA( $p, q$ ), e após a estimação dos seus parâmetros, o estimador dos resíduos (em cada instante de tempo  $t$ ) é dado por

$$\hat{\epsilon}_t = \hat{\Phi}(B)[\hat{\Theta}(B)]^{-1}(X_t - \hat{\mu}),$$

com  $\hat{\Phi}(B)$  e  $\hat{\Theta}(B)$ , respectivamente, os estimadores dos polinómios autoregressivos e médias móveis.

Pelo que podemos dizer que as estimativas dos resíduos se obtêm do seguinte modo:

$$\tilde{\epsilon}_t = X_t - \tilde{X}_t$$

com  $\tilde{X}_t$  o valor ajustado através do modelo ARMA( $p, q$ ), no instante  $t$ .

Ou seja, deve verificar-se se a série  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  verifica as seguintes propriedades:

1. se é uma série de ruído branco;
2. se segue uma distribuição normal e
3. se têm média nula e variância constante.

De um modo muito rápido, para verificar se os resíduos são ruído branco basta analisar as FAC e FACP amostrais dos resíduos estimados. Se quase todos os seus valores estiverem dentro dos respectivos intervalos de confiança então podemos afirmar, com uma confiança de  $(1 - \alpha)100\%$ , que os resíduos têm um comportamento semelhante ao de uma série de ruído branco. Tal como foi referido na secção 3.2.3, alguns valores podem ficar fora do intervalo de confiança.

Para verificar se os resíduos seguem uma distribuição normal o mais adequado é analisar os resíduos padronizados  $\hat{\epsilon}_t/S_\epsilon^2$  através da análise do cronograma, de um histograma (Pacheco, 2001) ou através de um teste de ajustamento, de Kolmogorov-Smirnov ou de Shapiro-Wilk, por exemplo, a  $\hat{\epsilon}_t$  ou aos dados padronizados.

No caso de se rejeitar a hipótese de normalidade dos resíduos, e se estes não se afastarem muito da normalidade, então todos os testes estatísticos aplicados aos resíduos, bem como as técnicas de previsão, continuam válidos.

De modo a verificar se os resíduos têm média nula basta realizar o usual teste t para a média ( $H_0 : \mu = 0$  vs  $H_1 : \mu \neq 0$ ). Para se verificar se os resíduos têm variância constante não é tão simples pois não existe nenhum teste estatístico. Por exemplo, *pode subdividir-se a série dos resíduos e calcular-se a variância amostral dos resíduos em cada um desses intervalos. Sempre que as variâncias amostrais, obtidas nesses intervalos, diferirem bastante entre si deverá procurar-se modelos alternativos* (Gomes, 2005).

Existem também dois tipos de testes estatísticos para testar se a série dos resíduos é um ruído branco: testes de hipótese sobre os valores individuais das FAC e FACP e testes de hipóteses globais.

Os testes de hipóteses sobre os valores individuais para as FAC e FACP são, respectivamente,

$$H_0 : \rho_\epsilon(k) = 0 \quad \text{vs} \quad H_1 : \rho_\epsilon(k) \neq 0, \quad \text{para cada } k \in \mathbb{N},$$

e

$$H_0 : \phi_\epsilon(k, k) = 0 \quad \text{vs} \quad H_1 : \phi_\epsilon(k, k) \neq 0, \quad \text{para cada } k \in \mathbb{N},$$

O primeiro designa-se por teste de Bartlett e o segundo por teste de Jenkins e Daniels (ver mais pormenores em (Pacheco, 2001) e (Gomes, 2005)).

O teste global testa a nulidade de uma sequência de  $m$  valores iniciais da FAC. Ou seja, testa

$$H_0 : \rho_\epsilon(1) = \rho_\epsilon(2) = \dots = \rho_\epsilon(m) = 0 \quad \text{vs} \quad H_1 : \exists \rho_\epsilon(k) \neq 0, \quad \text{para } k = 1, 2, \dots, m,$$

Este teste de hipóteses designa-se por teste de Box-Pierce (ou teste de “Portemanteau”) ou por Ljung e Box, dependendo da estatística de teste que se aplique (Gomes, 2005).

Existem ainda *testes de aleatoriedade dos resíduos que pretendem averiguar se os valores observados dos resíduos são provenientes de uma série de variáveis aleatórias independentes e identicamente distribuídas com distribuição absolutamente contínua*. São eles o teste do sinal das diferenças (*Difference-Sign Test*), o teste dos instantes de inflexão (*Turning Points Test*) e o teste das ordens (*Rank Test*) (Pacheco, 2001).

### 3.5.4 Critérios de selecção de modelos

Pode acontecer que mais do que um modelo se mostre adequado para descrever os dados e todos eles cumpram os pressupostos anteriores. Neste caso é preciso escolher o “melhor”deles.

Existem quatro critérios usuais:

- AIC — critério de informação de Akaike
- BIC — critério Bayesiano de informação de Akaike
- SBC — critério Bayesiano de Schwartz
- AICC — critério de informação de Akaike corrigido

Quanto menores forem os valores destes critérios mais adequados são os modelos para descrever os dados. Mais tarde alguns autores provaram que o valor do AIC diminuía à medida que o valor de parâmetros do modelo aumentava. Assim, surgiram o critério de informação de Akaike corrigido e o critério Bayesiano de informação de Akaike que não são influenciados pelo número de parâmetros no modelo (Clement, 2014).

Em geral, os critérios AIC e AICC tentam encontrar o melhor modelo que descreve o conjunto das observações, enquanto o critério BIC tenta encontrar o melhor ajustamento.

**Nota:** Em modelos não estacionários, os critérios de comparação, só poderão ser utilizados para procurar o melhor modelo de entre os vários modelos propostos dentro das mesmas transformações para a transformar a série temporal em estacionária (Gomes, 2005).

Para diferentes transformações deve utilizar-se outros critérios de comparação, como os que iremos apresentar no caso da avaliação das previsões.

### 3.6 Previsão de séries temporais

Um dos objectivos do modelo ajustado é utilizar-lo para fazer previsões de valores futuros. E a qualidade dessas previsões também é um bom indicador sobre a qualidade do modelo ajustado.

Sendo assim se  $t$  é o período corrente estamos interessados em prever os valores de  $X_{t+1}, X_{t+2}, \dots$

Prova-se (Murteira, 1993) que o melhor preditor de  $X_{t+k}$  para  $k = 1, 2, \dots$ , em termos do erro quadrático médio, é a esperança condicional que se representa usualmente como

$$X_t(k) = E[X_{t+k} | X_t, X_{t-1}, \dots]$$

ao valor inteiro  $k$  chama-se horizonte de previsão.

**Exemplo 3.6.1.** *Seja  $X_t$  um processo AR(1) de média não nula. Então, é possível obter a previsão de  $X_{t+k}$ , conhecida a série  $(X_1, X_2, \dots, X_t)$ , de modo recursivo, do seguinte modo*

$$X_t(1) = E[X_{t+1} | X_1, \dots, X_t]$$

$$= \hat{\mu} + \hat{\phi}X_t$$

*dado que  $\epsilon_{t+1}$  é independente de  $(X_1, X_2, \dots, X_t)$  e, portanto,  $E[\epsilon_{t+1} | X_1, \dots, X_t] = E[\epsilon_{t+1}] = 0$ .*

$$X_t(2) = E[X_{t+2} | X_1, \dots, X_t]$$

$$= \hat{\mu} + \hat{\phi}X_t(1)$$

*e assim sucessivamente.*

O erro de previsão em  $k$  passos é definido como

$$e_t(k) = X_{t+k} - X_t(k)$$

e prova-se que

$$E[e_t(k)] = 0$$

e que para além disso, através do cálculo da  $Var[e_t(k)]$ , é possível obter-se intervalos de confiança para as previsões. Se os processos forem não estacionários também é possível obterem-se previsões. (Pacheco, 2001).

### 3.6.1 Performance preditiva

Um das práticas muito utilizada para avaliar a qualidade das previsão dentro do período amostral consiste em retirar uma parte da série (em geral os últimos dados) e modelar a série temporal sem esta informação. Chama-se a este conjunto de dados que se deixa de fora a série de teste. Depois realizam-se previsões para esse horizonte que se retirou e os resultados previstos são comparados com os reais valores observados.

Seja  $\{\epsilon_t(1), t = k, \dots, n-1, (k \geq 1)$  os erros de previsão a um passo relativos à série  $(X_1, X_2, \dots, X_t)$ .

No momento da escolha do modelo, caso houver dúvida na escolha do modelo com maior performance preditiva, podemos considerar os seguintes três critérios baseados nos erros de previsão é um passo para ajudar na seleção do melhor modelo (Gomes, 2005):

- Critério do Erro Quadrático Médio(EQM)

$$EQM = \frac{1}{n-k} \sum_{t=k}^{n-1} [\epsilon_t(1)]^2$$

- Critério do Erro Absoluto Médio(EAM)

$$EAM = \frac{1}{n-k} \sum_{t=k}^{n-1} |\epsilon_t(1)|$$

- Critério do Erro Percentual Absoluto Médio(EPAM)

$$EPAM = \frac{1}{n-k} \sum_{t=k}^{n-1} \left| \frac{\epsilon_t(1)}{X_t} \right| * 100$$

Segundo estes critérios, o modelo mais adequado será aquele cujo os erros apresentem valores mais baixos.

# 4

## Aplicação ao caso de estudo

---

4.1	Introdução . . . . .	54
4.2	Tratamento prévio dos dados . . . . .	54
4.3	Análise descritiva dos dados . . . . .	54
4.4	Modelação das séries . . . . .	58
4.5	Beneficiário da política de controlo do preço de combustível 2001-2017 . . . . .	71

---

## 4.1 Introdução

Depois de caracterizado o mercado dos combustíveis e descrito a empresa ENCO, SARL, assim como depois de estudadas as metodologias de Séries Temporais iremos, neste capítulo, fazer uma aplicação de estas ao caso de estudo. Mais concretamente, ao estudo das séries dos diferenciais de preço da gasolina, do gasóleo e do petróleo entre os anos de 2001 e 2017.

Antes da aplicação das metodologias de Séries Temporais aos dados foi necessário proceder a um tratamento prévio dos dados, que apresentaremos na secção seguinte. Depois de os dados estarem adequados à aplicação das metodologias foram, então, ajustados modelos SARIMA às três séries de dados e, posteriormente, procedeu-se à previsão de valores futuros através dos modelos ajustados.

E, finalmente, enunciam-se as principais vantagens e desvantagens da utilização dos métodos de previsão aplicados, propondo melhorias futuras.

## 4.2 Tratamento prévio dos dados

Antes da aplicação dos modelos de Séries Temporais às três séries dos diferenciais de preços foi necessário proceder a um tratamento prévio dos dados. Em primeiro lugar, os dados originais não eram igualmente espaçados, o que impossibilitaria o ajustamento dos modelos ARMA/SARIMA às séries observadas. Em segundo lugar, entre os anos de 2001 a 2004 os valores observados variam em torno do zero, a semelhança do ruído branco, originado pela variância o preço real que era muito baixa neste período, como se pode observar nos gráficos seguintes. Devido a esta diferença de comportamento entre os períodos de 2001 a 2004 e de 2005 a 2017, foi apenas considerado para este estudo o período entre 2005 a 2017, uma vez que se considerássemos todo o período em causa isso iria influenciar, de modo negativo, os modelos a ajustar e as consequentes previsões.

De modo a ultrapassar o facto de os dados não serem igualmente espaçados, criaram-se médias intermédia nos anos onde faltavam dados, para que todos os anos tivessem 12 entradas, fazendo com que as séries fossem mensais. Esta alteração não causou grande alteração na estrutura da série original. Em anexo (6.1, 6.2, 6.3) ilustra-se os gráficos das séries originais.

## 4.3 Análise descritiva dos dados

Tal como foi referido no Capítulo 3, a realização da análise gráfica aos dados tem um papel extremamente importante para o ajustamento dos modelos. A análise gráfica permite a identificação de padrões, ou observações de casos pontuais, fazendo com que os mesmos sejam tomados em consideração na escolha do modelo.

Os gráficos seguintes dizem respeito aos valores do diferencial de preço (em dobras), ao longo dos 12 meses do ano, da gasolina, do gasóleo e do petróleo entre os anos de 2001 e 2017, que a ENCO, SARL calcula com base no mapa de estrutura de preço dos combustível.

Todas as análises foram feitas através do *software* estatístico “R”.

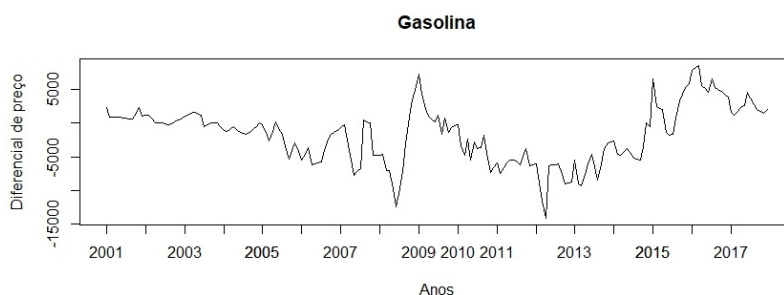
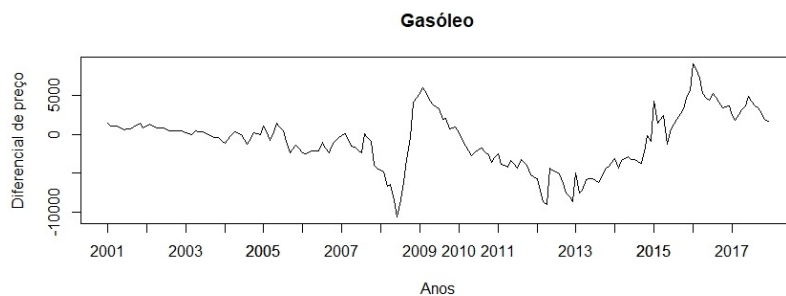
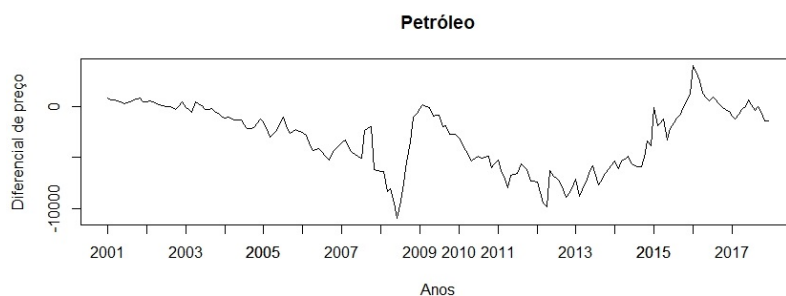


Figura 4.1: Cronograma dos diferenciais de preço da gasolina (dobras/mês)



**Figura 4.2:** Cronograma dos diferenciais de preço do gasóleo (dobras/mês)



**Figura 4.3:** Cronograma dos diferenciais de preço do petróleo (dobras/mês)

Como se pode verificar pela análise dos gráficos de gasolina, do gasóleo e do petróleo entre os anos de 2001 a 2004 os valores variam em torno do zero a semelhança do ruído branco. Devido a esse comportamento os valores relativos a esses anos não foram considerados neste caso de estudos, já que iriam influenciar no resultado dos modelos a ajustar e às consequentes previsões, tal como referido anteriormente.

Os gráficos 4.4—4.6 dizem respeito às séries dos combustíveis (gasolina, gasóleo e petróleo), após serem retirados os anos acima mencionados. Estes gráficos foram obtidos através do método de decomposição clássica (do tipo aditivo) de séries temporais, que consiste em estimar, em separado, as componentes tendência, sazonalidade e erros (Ehlers, 2007).

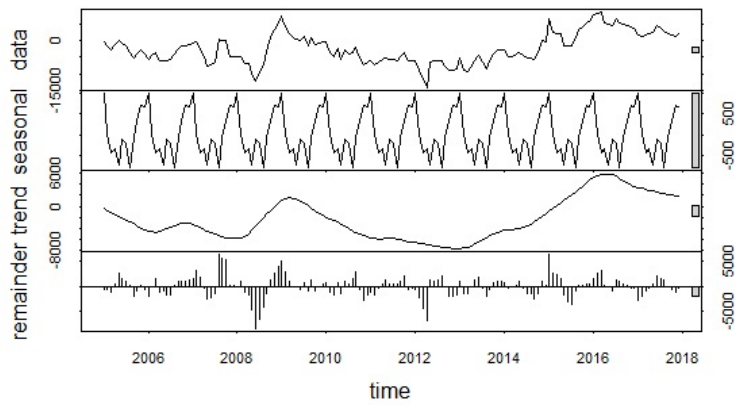
Tal como se pode verificar, aparentam ser séries não-estacionárias, porque à simples vista nota-se que as séries apresentam características de tendência.

Entre os anos de 2005 e final de 2007 existe uma tendência decrescente em todas as séries, mas com pequenas ondulações. No ano 2008, com a crise internacional, houve um aumento considerável do preço de produtos petrolíferos a nível internacional o que originou um aumento acentuado do diferencial de preço. Por não se ter ajustado o preço de venda ao consumidor final, em mediados de 2009 os preços voltaram a descer até 2013, onde retomou a sua tendência crescente até o princípio do ano 2016.

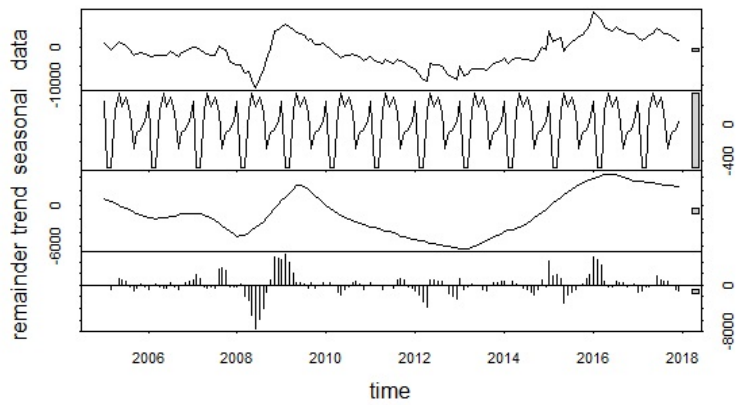
Conforme observado nos gráficos das séries temporais, a componente de tendência dos gráficos de decomposição da gasolina, do gasóleo e do petróleo confirmam a existência de componente tendencial nas séries, logo as séries são não-estacionárias na parte tendencial, ou seja, as séries aparentam ser não-estacionária na média.

Através dos gráficos das séries e dos gráficos de decomposição não foi possível detectar componentes sazonais. Para uma análise um pouco mais profunda, vai agora averiguar-se com melhor precisão os gráficos dos diferenciais em separado para cada ano.

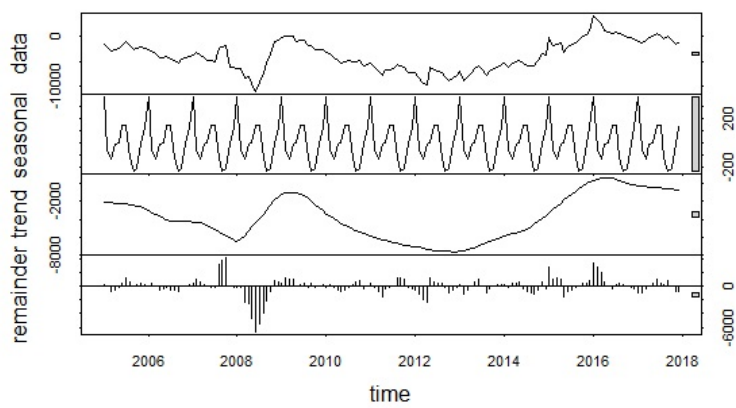
Os gráficos acima ilustram que, tanto a gasolina, como o gasóleo e o petróleo, não apresentam repetições sistematicamente ano após ano. Portanto, não foi possível distinguir padrões de sazonalidade nestes dados.



**Figura 4.4:** Decomposição dos diferenciais de preço da gasolina

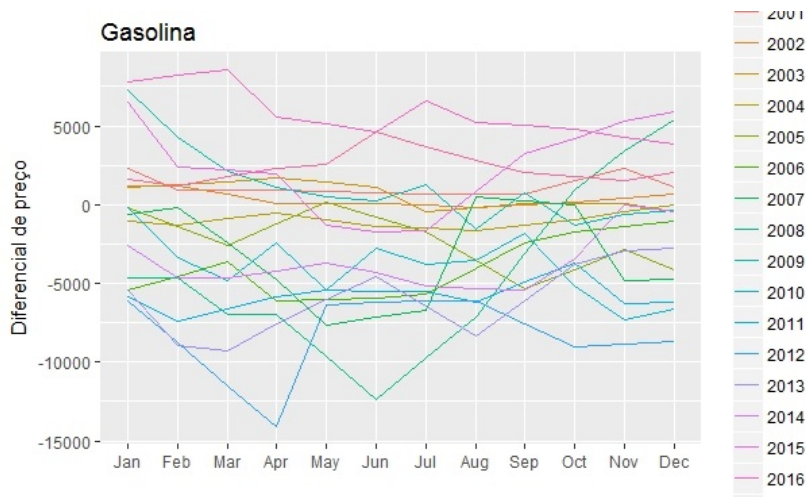


**Figura 4.5:** Decomposição dos diferenciais de preço do gasóleo

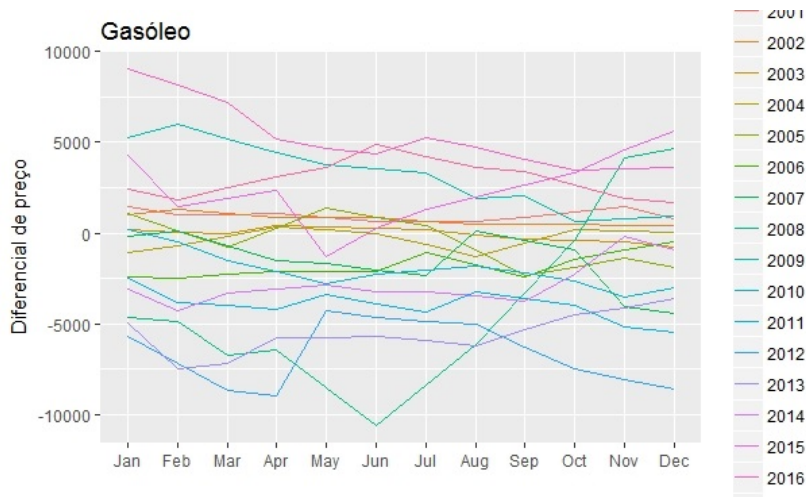


**Figura 4.6:** Decomposição dos diferenciais de preço do petróleo

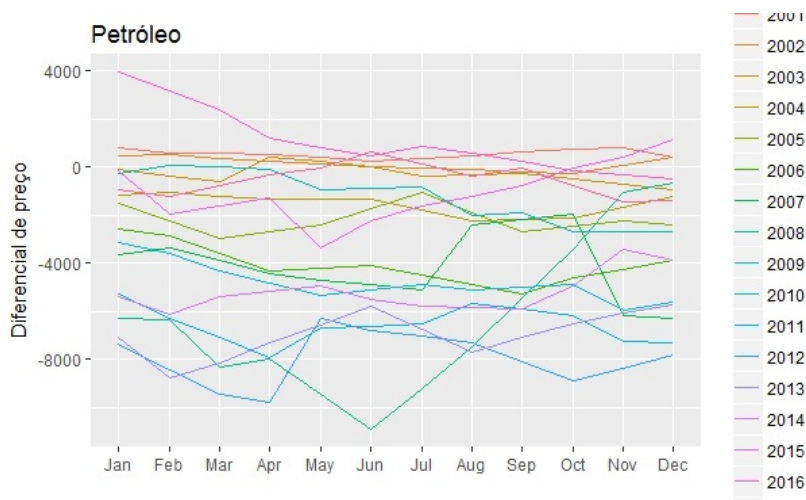




**Figura 4.7:** Gráfico dos diferenciais de preço da gasolina (dobras/mês) por ano



**Figura 4.8:** Gráfico dos diferenciais de preço do gasóleo (dobras/mês) por ano



**Figura 4.9:** Gráfico dos diferenciais de preço do petróleo (dobras/mês) por ano

Após a análise gráfica, que nos deu a possibilidade de conhecer as nossas séries temporais, iremos ainda fazer uma análise descritiva dos dados observados.

Através do cálculo de algumas medidas de dispersão (desvio padrão e amplitude interquartil) é possível ter uma noção da variabilidade dos dados. Quanto maior forem estes valores mais dispersos são os dados. Entretanto, para uma melhor compreensão da análise da dispersão, elaborou-se a tabela apresentada na Figura 4.10, que contém os resultados da análise descritiva, tanto anual como para toda a série, para todos os combustíveis.

Se consideramos apenas o desvio padrão como a medida para avaliar a dispersão, o produto que apresenta uma maior dispersão de valores é a gasolina. Ao realizar a razão entre o desvio padrão e a média (coeficiente de variação - CV) apura-se que é o que o petróleo aquele que aparenta maior homogeneidade, com o valor CV de aproximadamente 2,44. O gasóleo é o produto que apresenta o maior valor do CV.

Através da medida de dispersão alternativa ao desvio padrão, a distância interquartil, que é mais conservadora que o desvio padrão (ou seja, menos sensível a valores extremos), a gasolina é o produto que apresenta maior dispersão dos dados com valor  $\approx 6.157,24$ , em seguida o petróleo com valor  $\approx 5.223,92$  e o produto que apresenta os dados mais homogêneos é o gasóleo com valor  $\approx 4.299,34$ .

Análise numérica em conjunto com a análise gráfica, ajuda-nos a ter mais informações sobre a série e na escolha do modelo que melhor adequam as características dos dados da série.

Ano		Gasolina	Gasóleo	Petróleo	Ano		Gasolina	Gasóleo	Petróleo	Ano		Gasolina	Gasóleo	Petróleo
2	Média	1.137,17	962,87	550,45	2	Média	354,29	738,41	145,75	2	Média	602,40	-56,54	-286,67
0	Mediana	935,75	949,50	554,75	0	Mediana	117,50	747,25	108,00	0	Mediana	585,75	9,25	-319,25
0	Desvio padrão	602,32	284,99	183,54	0	Desvio padrão	471,48	297,66	258,86	0	Desvio padrão	797,92	375,71	390,78
0	1. Quartil inferior	737,00	758,00	399,00	0	1. Quartil inferior	53,38	476,87	-22,12	0	1. Quartil inferior	34,88	-369,00	-514,63
1	3. Quartil superior	1.177,50	1.078,37	673,00	2	3. Quartil superior	658,13	901,12	391,12	3	3. Quartil superior	1.293,19	207,75	-77,50
	Amplitude interquart	440,50	320,37	274,00		Amplitude interquartil	604,75	424,25	413,24		Amplitude interquartil	1.258,31	576,75	437,13
2	Média	-992,94	-309,89	-1.548,37	2	Média	-2.316,88	-423,58	-2.188,37	2	Média	-3.985,20	-1.805,81	-4.081,42
0	Mediana	-976,50	-97,25	-1.335,75	2	Mediana	-2.163,50	-335,50	-2.317,00	2	Mediana	-4.290,30	-2.106,50	-4.222,75
0	Desvio padrão	478,00	538,83	424,27	0	Desvio padrão	1.728,70	1.278,20	549,81	0	Desvio padrão	1.934,97	664,88	775,98
0	1. Quartil inferior	-1.315,25	-663,75	-1.863,62	0	1. Quartil inferior	-3.669,10	-1.521,50	-2.512,25	0	1. Quartil inferior	-5.733,50	-2.313,25	-4.522,00
4	3. Quartil superior	-816,62	112,82	-1.203,37	5	3. Quartil superior	-1.131,87	535,63	-1.835,87	6	3. Quartil superior	-2.226,75	-1.357,63	-3.811,13
	Amplitude interquart	498,63	776,57	660,25		Amplitude interquartil	2.537,23	2.057,13	676,38		Amplitude interquartil	3.506,75	955,63	710,88
2	Média	-3.192,20	-1.508,70	-4.071,80	2	Média	-4.622,30	-4.266,90	-6.386,90	2	Média	1.148,20	3.138,10	-1.243,90
0	Mediana	-3.594,80	-1.217,50	-4.145,50	2	Mediana	-5.821,00	-5.507,50	-6.917,50	2	Mediana	6.645,00	3.415,50	-921,00
0	Desvio padrão	3.132,50	1.515,00	1.450,40	0	Desvio padrão	5.451,60	4.834,50	3.242,30	0	Desvio padrão	2.506,40	1.869,40	1.100,30
0	1. Quartil inferior	-5.276,00	-2.124,30	-4.949,00	0	1. Quartil inferior	-7.744,50	-7.148,00	-8.539,50	0	1. Quartil inferior	-408,13	1.630,30	-2.163,00
7	3. Quartil superior	-115,50	-339,40	-3.095,30	8	3. Quartil superior	-2.095,60	-2.554,50	-4.949,13	9	3. Quartil superior	1.463,50	4.608,80	-229,00
	Amplitude interquart	5.160,50	1.784,90	1.853,70		Amplitude interquartil	5.648,90	4.993,50	3.590,37		Amplitude interquartil	1.871,63	2.978,50	1.934,00
2	Média	-3.938,40	-2.024,30	-4.797,14	2	Média	-5.783,04	-3.970,17	-6.547,34	2	Média	-8.285,30	-6.653,80	-7.952,20
0	Mediana	-3.694,50	-2.185,30	-4.927,75	2	Mediana	-5.850,29	-3.932,19	-6.552,52	2	Mediana	-8.130,70	-6.731,20	-7.943,70
0	Desvio padrão	2.052,10	1.052,80	798,46	0	Desvio padrão	911,14	798,78	755,64	0	Desvio padrão	2.496,20	1.724,00	1.070,70
1	1. Quartil inferior	-5.266,50	-2.674,00	-5.154,38	1	1. Quartil inferior	-6.190,39	-4.236,28	-7.123,49	1	1. Quartil inferior	-8.902,40	-8.201,10	-8.520,70
0	3. Quartil superior	-2.646,00	-1.735,75	-4.687,38	1	3. Quartil superior	-5.450,39	-3.576,62	-6.109,67	2	3. Quartil superior	-6.192,60	-4.998,50	-7.225,00
	Amplitude interquart	2.620,50	938,25	467,00		Amplitude interquartil	740,00	659,66	1.013,82		Amplitude interquartil	2.709,80	3.202,60	1.295,70
2	Média	-6.015,70	-5.542,30	-6.949,39	2	Média	-3.668,40	-2.812,05	-5.173,72	2	Média	2.331,70	2.364,60	-1.046,10
0	Mediana	-6.069,30	-5.699,50	-6.903,76	2	Mediana	-4.259,10	-3.147,90	-5.361,31	2	Mediana	2.297,90	2.167,90	-1.239,80
0	Desvio padrão	2.238,90	1.144,70	925,81	0	Desvio padrão	1.796,40	1.171,30	818,02	0	Desvio padrão	2.886,50	1.908,30	1.241,80
1	1. Quartil inferior	-7.765,60	-5.992,30	-7.412,81	1	1. Quartil inferior	-4.800,30	-3.355,40	-5.799,45	1	1. Quartil inferior	279,73	1.395,85	-1.708,20
3	3. Quartil superior	-4.362,40	-4.852,20	-6.384,48	4	3. Quartil superior	-3.262,70	-2.727,25	-4.922,51	5	3. Quartil superior	4.509,40	3.534,00	-127,76
	Amplitude interquart	3.403,20	1.140,10	1.028,33		Amplitude interquartil	1.537,60	628,15	876,94		Amplitude interquartil	4.229,67	2.138,15	1.580,44
2	Média	5.803,50	5.267,03	1.065,13	2	Média	2.320,29	2.986,36	-547,56	2	Média	-1.711,94	-818,63	-3.003,50
0	Mediana	5.194,20	4.672,64	7.159,65	2	Mediana	2.038,55	2.881,39	-561,56	0	Mediana	-1.262,50	-400,50	-2.314,50
0	Desvio padrão	1.602,50	1.874,20	1.420,23	0	Desvio padrão	973,77	990,93	654,80	0	Desvio padrão	4.190,50	3.543,29	3.052,60
1	1. Quartil inferior	4.727,65	3.964,35	126,08	1	1. Quartil inferior	1.706,95	2.314,99	-1.017,12	1	1. Quartil inferior	-5.166,86	-3.223,71	-5.516,80
6	3. Quartil superior	6.889,67	5.731,74	1.510,92	7	3. Quartil superior	2.608,71	3.595,23	-15,69	1	3. Quartil superior	990,38	1.075,63	-292,88
	Amplitude interquart	2.162,02	1.767,39	1.384,84		Amplitude interquartil	901,77	1.280,24	1.001,43	7	Amplitude interquartil	6.157,24	4.299,34	5.223,92

Figura 4.10: Análise descritiva dos diferenciais de preço da gasolina, gasóleo e petróleo, por ano

Após a preparação dos dados e à sua análise descritiva iremos, então, passar à fase seguinte do ajustamento dos modelos, ou modelação das séries dos dados.

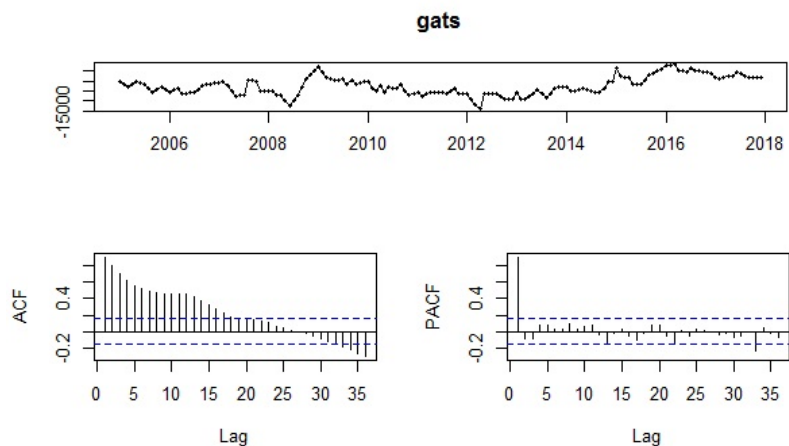
Todas as análises inferenciais foram feitas considerando o nível usual de significância de 5%.

## 4.4 Modelação das séries

### 4.4.1 Modelação dos diferenciais de preço da gasolina

Como já foi abordado no capítulo 3, o primeiro passo para identificação do(s) modelo(s) adequados para modelar uma série temporal é observar as suas FAC e FACP.

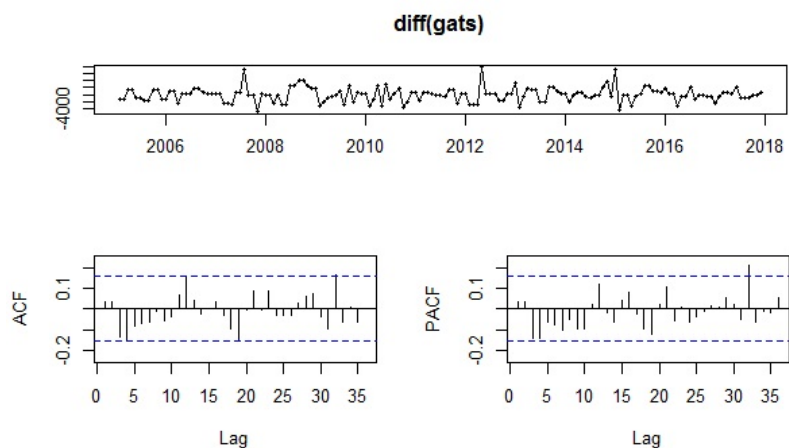
Portanto, de seguida apresenta-se a série dos diferenciais de preço da gasolina e as respectivas FAC e FACP empíricas.



**Figura 4.11:** FAC e FACP dos diferenciais de preço da gasolina (gats)

Através do gráfico, podemos uma vez mais comprovar a não estacionariedade da série, uma vez que as FAC tendem muito lentamente para zero, o que representa um forte indício que a série é não estacionária, tal como referido no Capítulo 3.

Sendo, então, a série não estacionária transformou-se esta em uma série estacionária, através das transformações às séries também referidas no Capítulo 3. Ou seja, começou por verificar-se a necessidade de aplicar-se uma transformação de Box-Cox. Obteve-se um valor estimado de  $\lambda$  igual a 1, pelo que não foi necessário fazer esta transformação. De seguida, aplicou-se então uma diferenciação simples, obtendo-se o seguinte resultado



**Figura 4.12:** Série transformada dos diferenciais de preço da gasolina (diff(gats))

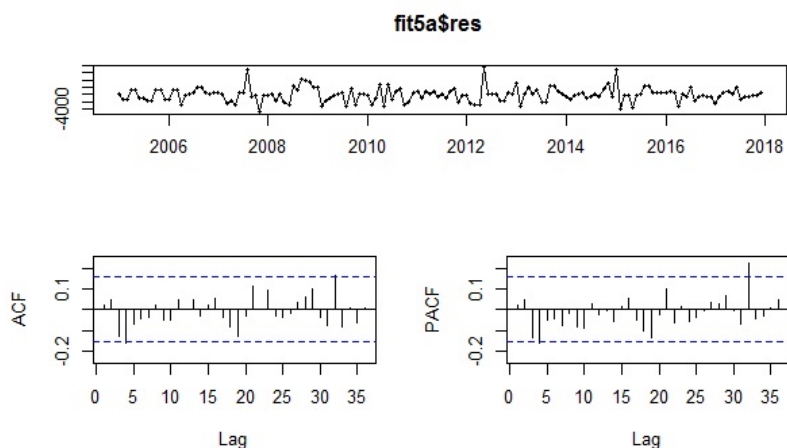
Note agora que depois da diferenciação, temos uma série estacionária. Ao analisarmos as FAC e FACP a questão seguinte é qual a estrutura do(s) modelo(s) do tipo ARMA e quais as suas ordens.

Depois do exposto, considerou-se dois possíveis modelos que podem ser adequados para descrever a série original são:  $SARIMA(0, 1, 0) \times (0, 0, 1)_{12}$  e  $SARIMA(0, 1, 0) \times (1, 0, 0)_{12}$ .

Após a escolha dos possíveis modelos foram, então, estimados os seus parâmetros.

De seguida procedeu-se à significância dos parâmetros estimados, através dos respectivos intervalos de confiança para os parâmetros dos modelos.

Seguidamente, e após a estimação e avaliação da significância dos parâmetros dos modelos  $SARIMA(0, 1, 0) \times (0, 0, 1)_{12}$ , vai proceder-se à análise dos resíduos de modo a avaliar a adequação do modelo.



**Figura 4.13:** Resíduos dos diferenciais de preço da gasolina (modelo  $SARIMA(0, 1, 0) \times (0, 0, 1)_{12}$ )

Ao analisar os resíduos, a sua série e as respectivas FAC e FACP, tendo em conta as características que devem ter os resíduos (ver subsecção 3.5.3.E), os mesmos são aceitável. Neste sentido, não se rejeita a hipótese de os resíduos se comportarem como um ruído branco.

#### Testes aos resíduos do modelo $SARIMA(0, 1, 0) \times (0, 0, 1)_{12}$

Para o nível de significância de 5%.

- Para verificar se os resíduos são provenientes de uma população com distribuição normal, realizou-se também o teste de Kolmogorov-Smirnov para testa a normalidade dos resíduos, e obteve-se um *p-value* de 0,156. Pelo que não se rejeita a hipótese nula da normalidade dos dados para o nível de significância de 5%.
- Também se realizou um teste t-student, para verificar se os resíduos são provenientes de uma população com média nula, tendo-se obtido um *p-value* de 0,942. Pelo que não se rejeita a hipótese nula da nulidade da média dos resíduos, para o nível de significância de 5%.
- De forma a testar a aleatoriedade dos resíduos, realizou-se o teste da diferença de sinais (Difference-Sign Test), e obteve-se um *p-value* de 0,489. Pelo que não se rejeita a hipótese nula para o nível de significância de 5%.

Realizou-se os mesmos testes aos resíduos do modelo  $SARIMA(0, 1, 0) \times (1, 0, 0)_{12}$ , e verificou-se os mesmos pressupostos, ou seja, não se rejeita a hipótese nula da normalidade dos dados para o nível de significância de 5%.

Tendo em conta os resultado dos teste acima abordados, pode-se afirma de uma forma geral que as características dos resíduos de ambos os modelos são aceitáveis. Neste caso, como ambos modelos estão bem, vai-se ver qual deles é o melhor.

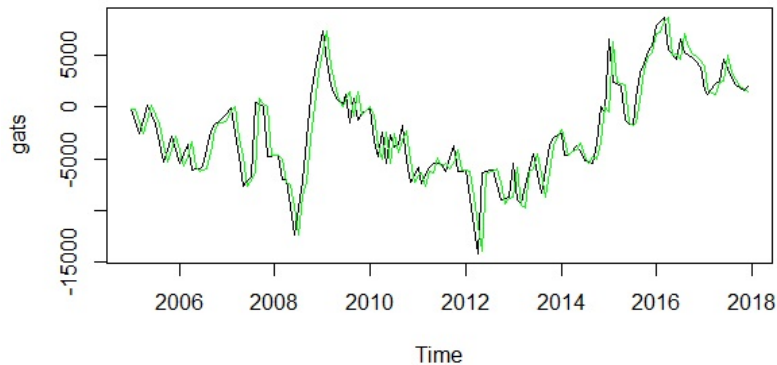
Considerando que os testes acima realizados dão como aptos os dois modelos propostos, neste caso temos que escolher o melhor dos dois. Portanto vai utilizar-se os critérios da subsecção 3.5.4 para identificar o melhor modelo.

Após a comparação dos resultados dos critérios AIC, AICC e BIC dos dois modelos, e tendo em conta que ambos têm o mesmo número de parâmetros (princípio da parcimónia), o modelo  $SARIMA(0, 1, 0) \times (0, 0, 1)_{12}$  é o que apresenta os menores valores e, portanto, é o modelo que se vai considerar como o melhor para descrever os dados.

De seguida apresentamos o gráfico do ajustamento deste modelo aos dados originais

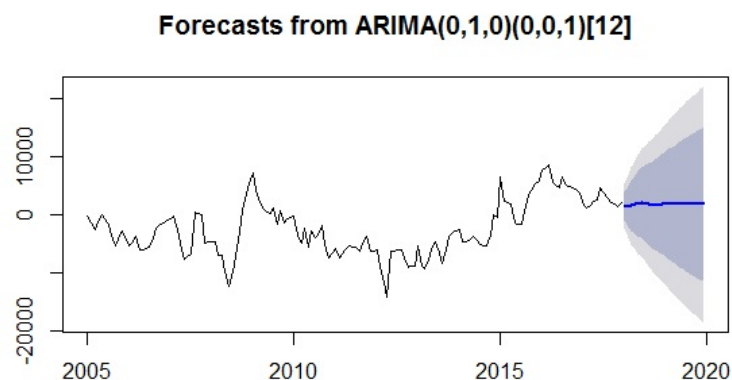
Modelo	AIC	AICc	BIC
SARIMA(0, 1, 0) × (0, 0, 1) <sub>12</sub>	2789.82	2789.9	2795.9
SARIMA(0, 1, 0) × (1, 0, 0) <sub>12</sub>	2790.24	2790.32	2796.32

**Tabela 4.1:** Critérios de comparação relativos ao ajustamento dos diferenciais de preço da gasolina



**Figura 4.14:** Ajustamento do modelo SARIMA(0, 1, 0) × (0, 0, 1)<sub>12</sub> aos diferenciais de preço da gasolina. Dados originais (a preto) e ajustados (a verde)

De acordo com o gráfico acima, pode ver-se que o modelo consegue descrever bem a evolução temporal da série da gasolina. Contudo, tal como foi referido no Capítulo 3, a qualidade das previsões também é um bom indicador sobre a qualidade do modelo ajustado. Pelo que vamos agora fazer previsões através do modelo SARIMA(0, 1, 0) × (0, 0, 1)<sub>12</sub> ajustado.



**Figura 4.15:** Previsão dos diferenciais de preço da gasolina para 2018-2019 (a azul)

Através do gráfico pode ver-se as previsões para os anos de 2018 e 2019 e as respectivas banda de confiança a 80% e 95% (ver subsecção 3.6).

Foram Também calculadas as medidas de EQM, EAM e EPAM para avaliar a performance preditiva dos modelos (ver subsecção 3.6.1).

	gats.real	a23\$mean	fita.prev23a
Jan 2017	1576.70	3957.623	5654.877
Feb 2017	1199.06	4155.707	5662.265
Mar 2017	1750.36	4232.354	4781.587
Apr 2017	2301.65	3683.868	6883.410
May 2017	2548.46	3723.636	6126.889
Jun 2017	4581.31	3636.446	4059.381
Jul 2017	3685.39	3988.262	6974.957
Aug 2017	2789.47	3668.303	4767.863
Sep 2017	2059.38	3539.501	7627.698
Oct 2017	1797.79	3471.841	10019.251
Nov 2017	1536.19	3370.416	12328.650
Dec 2017	2017.71	3266.045	11503.209

**Figura 4.16:** Dados reais (gats.real) e previsão da gasolina para o ano de 2017 modelo

$$\text{SARIMA}(0, 1, 0) \times (0, 0, 1)_{12}$$

Erros de previsão do gasolina			
Modelos	EQM	EAM	EPAM
SARIMA (0,1,0) (0,0,1)12	1925.89	1381.829	248.2021
SARIMA (0,1,0)(1,0,0) 12	1928.952	1382.289	236.4412

**Figura 4.17:** Cálculo dos erros de previsão da gasolina

Tendo em conta os resultados dos erros de previsão calculados para ambos os modelos, o modelo  $\text{SARIMA}(0, 1, 0) \times (0, 0, 1)_{12}$  em dois erros (EQM e EAM) apresenta resultados inferior, enquanto que o modelo  $\text{SARIMA}(0, 1, 0) \times (1, 0, 0)_{12}$  só apresenta menor valor no EPAM, neste sentido, continua-se a considerar o modelo  $\text{SARIMA}(0, 1, 0) \times (0, 0, 1)_{12}$  como o melhor para ajustar e fazer previsão da série.

#### 4.4.2 Modelação dos diferenciais de preço do gasóleo

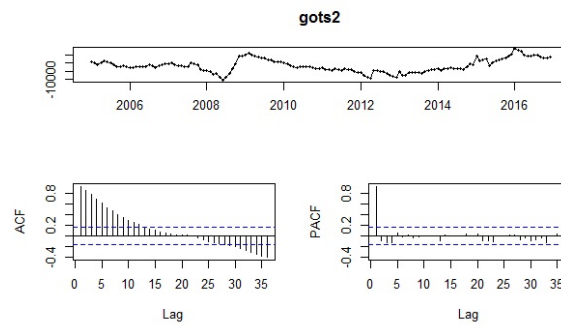
Ao contrário da série da gasolina, na série do gasóleo, de modo a avaliar a qualidade das previsões retirou-se o ano de 2017 (chamada série de teste) e ajustou-se um modelo à série dos diferenciais de preço do gasóleo entre os anos de 2005-2016.

Como já foi abordado na modelação da série anterior, o primeiro passo para identificação do(s) modelo(s) adequados para modelar uma série temporal é observar as suas FAC e FACP.

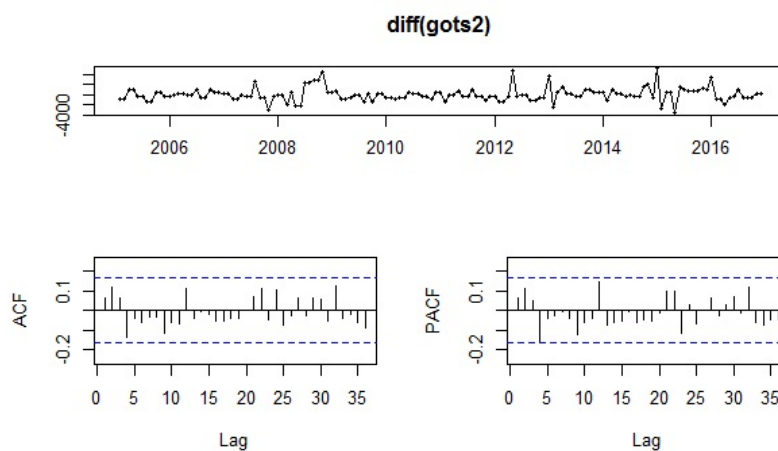
Portanto, de seguida apresenta-se a série teste dos diferenciais de preço da gasóleo e as respectivas FAC e FACP empíricas.

Através do gráfico, podemos uma vez mais comprovar a não estacionariedade da série, uma vez que as FAC tendem muito lentamente para zero, o que representa um forte indício que a série é não estacionária, tal como referido no Capítulo 3.

Sendo, então, a série não estacionária transformou-se esta em uma série estacionária, através das transformações às séries também referidas no Capítulo 3. Ou seja, começou por verificar-se a necessidade de aplicar-se uma transformação de Box-Cox. Obteve-se um valor estimado de  $\lambda$  igual a 1, pelo que não foi necessário fazer esta transformação. De seguida, aplicou-se então uma diferenciação simples, obtendo-se o seguinte resultado



**Figura 4.18:** FAC e FACP dos diferenciais de preço do gasóleo até 2016



**Figura 4.19:** Série dos diferenciais de preço do gasóleo transformada

Note agora que depois da diferenciação, temos uma série estacionária. Ao analisarmos as FAC e FACP a questão seguinte é qual a estrutura do(s) modelo(s) do tipo ARMA e quais as suas ordens.

Depois do exposto, considerou-se dois possíveis modelos que podem ser adequados para descrever a série original e que são:  $ARIMA(0, 1, 0)$  e  $ARIMA(1, 1, 0)$ .

Após a escolha dos possíveis modelos foram, então, estimados os seus parâmetros.

De seguida procedeu-se à significância dos parâmetros estimados, através dos respectivos intervalos de confiança para os parâmetros do modelo.

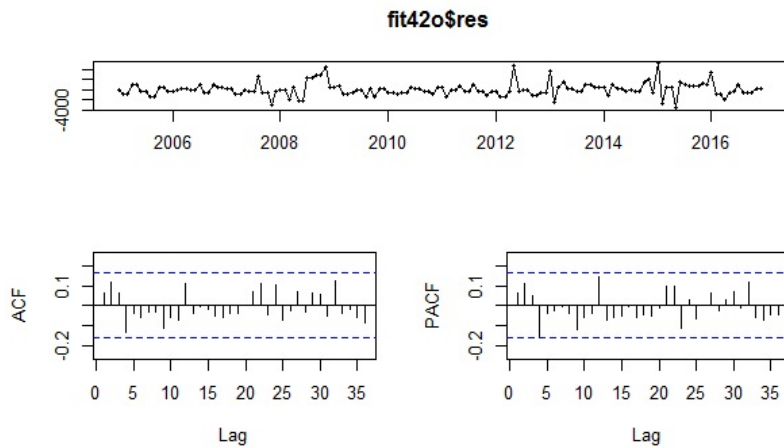
Seguidamente, e após a estimação e avaliação da significância dos parâmetros dos modelos, vai-se proceder à análise dos resíduos de modo a avaliar a adequação do modelo.

Ao analisar os resíduos, da série e as respectivas FAC e FACP, tendo em conta as características que devem ter os resíduos (ver subsecção 3.5.3.E), os mesmos são aceitáveis. Neste sentido, não se rejeita a hipótese de os resíduos se comportarem como um ruído branco.

#### Testes aos resíduos do modelo $ARIMA(0, 1, 0)$

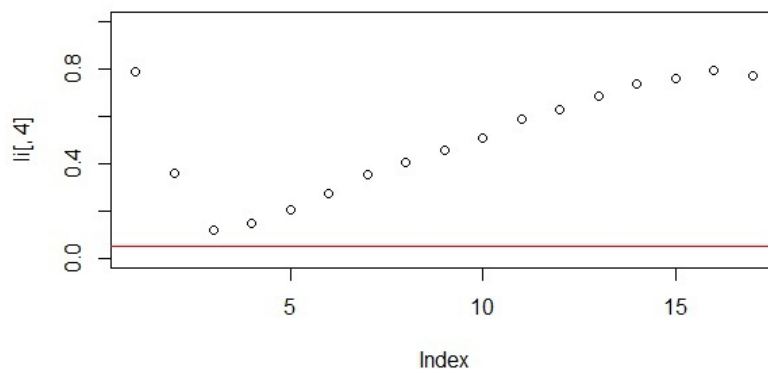
Para o nível de significância de 5%.

- Teste de Kolmogorov-Smirnov obteve um *p-value* de 0,015. Pelo que se rejeita a hipótese nula da normalidade dos dados.
- Teste t-student obteve um *p-value* de 0,872. Pelo que não se rejeita a hipótese nula da nulidade da média dos resíduos, para o nível de significância de 5%.



**Figura 4.20:** Resíduos do diferencial de preço do gasóleo modelo ARIMA(0,1,0)

- Teste da diferença de sinais (Difference-Sign Test) obteve um *p-value* de 0,0001. Pelo que se rejeita a hipótese nula, para o nível de significância de 5%.
- De forma a testar a linearidade na média, realizou-se o teste de Ljung-Box, e o resultado é aceitável, visto que não existe nenhum elemento fora da banda de confiança, como se pode comprovar na figura 4.23 abaixo.
- Também realizou-se o teste de LiMcLeod (ver anexo 6.7) para identificar a não linearidade na média ou na variância, e não existe nenhum elemento fora da banda de confiança, sendo assim o resultado é aceitável.



**Figura 4.21:** Teste Ljung-Box realizado a série do gasóleo modelo ARIMA(0,1,0) 2016

#### Testes aos resíduos do modelo ARIMA(1, 1, 0)

Para o nível de significância de 5%.

Também realizou-se os testes aos resíduos do modelo ARIMA(1, 1, 0) e obteve-se os seguintes resultados:

- Teste de Kolmogorov-Smirnov obteve um *p-value* de 0,039. Pelo que se rejeita a hipótese nula da normalidade dos dados.



- Teste t-student obteve um  $p$ -value de 0,88. Pelo que não se rejeita a hipótese nula da normalidade dos dados.
- Teste da diferença de sinais (Difference-Sign Test) obteve um  $p$ -value de 0,113. Pelo que não se rejeita a hipótese nula da normalidade dos dados.
- De forma a testa a linearidade na média, realizou-se o teste de Ljung-Box, e o resultado é aceitável, visto não existe nenhum elemento fora da banda de confiança, como se pode ver no anexo 6.9.
- Também realizou-se o teste de LiMcLeod (ver anexo 6.8) para identificar a não linearidade na média ou na variância, e não existe nenhum elemento fora da banda de confiança, sendo assim o resultado é aceitável.

Ao comparar os resultados dos testes acima abordados, não fica bem explícito qual dos dois teve melhores resultados. Portanto, vai-se utilizar os critérios da subsecção 3.5.4 para identificar o melhor modelo.

### Critérios de selecção de modelos

Após a comparação dos resultados dos critérios AIC, AICC e BIC dos dois modelos (tabela abaixo), e tendo em conta que ambos têm o mesmo número de parâmetros (princípio da parcimónia). Neste caso, o modelo  $ARIMA(0, 1, 0)$  é o que apresenta os menores valores e, portanto, é o modelo que se vai considerar como o melhor para descrever os dados.

Modelo	AIC	AICC	BIC
$ARIMA(0, 1, 0)$	2462.61	2462.64	2465.58
$ARIMA(1, 1, 0)$	2463.96	2464.04	2469.88

**Tabela 4.2:** Critérios de comparação relativos ao ajustamento dos diferenciais de preço do gasóleo

De modo a avaliar a qualidade das previsões retirou-se o ano de 2017 (chamada série de teste) e ajustou-se um modelo à série dos diferenciais de preço do gasóleo entre os anos de 2005-2016. De seguida apresentamos o gráfico do ajustamento deste modelo aos dados originais retirado 2017. No início, foi retirado da série o ano de 2017, para depois compara-lo com a previsão deste mesmo ano, de forma a comprovar a capacidade preditiva do modelo, e pode-se ver através do gráfico abaixo, que a previsão do modelo é aceitável.

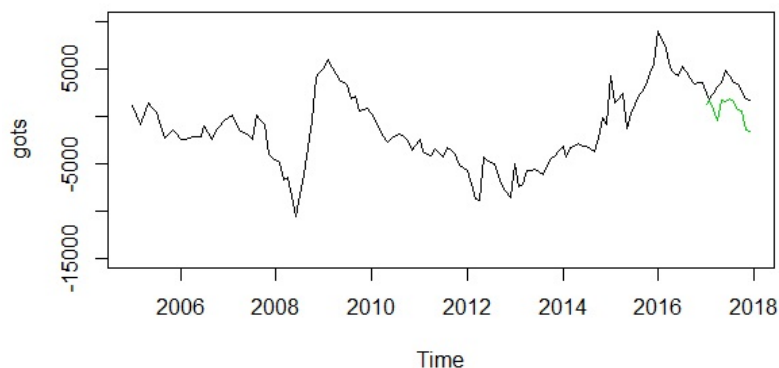
De seguida apresentamos o gráfico do ajustamento deste modelo aos dados originais

De acordo com o gráfico acima, pode-se ver que o modelo conseguiu descrever a série. Contudo, tal como foi referido no Capítulo 3, a qualidade das previsões também é um bom indicador sobre a qualidade do modelo ajustado, pelo que vamos agora fazer previsões do modelo  $ARIMA(0, 1, 0)$  ajustado.

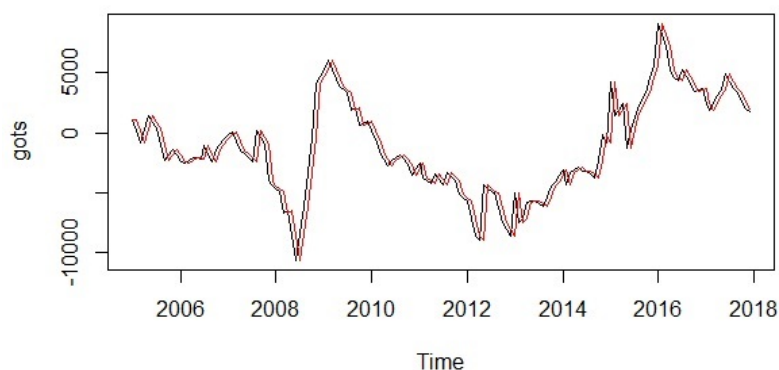
Através do gráfico, pode ver que as previsões para os anos de 2018 e 2019 e as respectivas bandas de confiança a 80% e 95% (ver subsecção 3.6).

Foram também, calculadas as medidas de EQM, EAM e EPAM para avaliar a performance preditiva dos modelos, ver subsecção 3.6.1.

Tendo em conta os resultados dos erros de previsão para ambos os modelos, o modelo  $ARIMA(0, 1, 0)$  em dois erros (EQM e EPAM) apresenta resultados inferior, enquanto que o modelo  $ARIMA(1, 1, 0)$  só apresenta menor valor no EAM.



**Figura 4.22:** Gráfico comparativo do real e previsão 2017 de gasóleo



**Figura 4.23:** Ajustamento do modelo  $ARIMA(0, 1, 0)$  aos diferenciais de preço do gasóleo. Dados originais (a preto) e ajustados (a vermelha)

### 4.4.3 Modelação dos diferenciais de preço do petróleo

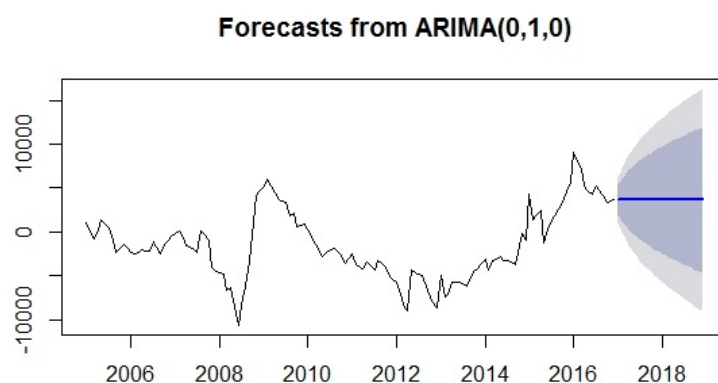
Usou-se o mesmo procedimento do gasóleo para a série do petróleo, ou seja, de modo a avaliar a qualidade das previsões retirou-se o ano de 2017 (chamada série de teste) e ajustou-se um modelo à série dos diferenciais de preço do gasóleo entre os anos de 2005-2016.

Como já foi abordado na modelação da série anterior, o primeiro passo para identificação do(s) modelo(s) adequados para modelar uma série temporal é observar as suas FAC e FACP.

Portanto, de seguida apresenta-se a série teste dos diferenciais de preço de petróleo e as respectivas FAC e FACP empíricas.

Através do gráfico, podemos uma vez mais comprovar a não estacionariedade da série, uma vez que as FAC tendem muito lentamente para zero, o que representa um forte indício que a série é não estacionária, tal como referido no Capítulo 3.

Sendo, então, a série não estacionária transformou-se esta em uma série estacionária, através das transformações às séries também referidas no Capítulo 3. Ou seja, começou por verificar-se a necessidade de aplicar-se uma transformação de Box-Cox. Obteve-se um valor estimado de  $\lambda$  igual a 1, pelo que não foi necessário fazer esta transformação. De seguida, aplicou-se então uma diferenciação simples, obtendo-se o seguinte resultado



**Figura 4.24:** Previsão dos diferenciais de preço da gasolina para 2018-2019 modelo ARIMA(0, 1, 0)

Data	Real	Mean	Previsão
Jan 2017	2445.77	3626.12	4312.452
Feb 2017	1834.92	3626.12	4564.460
Mar 2017	2471.54	3626.12	5439.343
Apr 2017	3108.16	3626.12	6880.263
May 2017	3583.69	3626.12	6388.329
Jun 2017	4856.08	3626.12	4672.384
Jul 2017	4242.96	3626.12	4614.733
Aug 2017	3629.84	3626.12	4219.240
Sep 2017	3386.61	3626.12	4614.773
Oct 2017	2654.62	3626.12	5662.882
Nov 2017	1922.63	3626.12	4540.006
Dec 2017	1699.53	3626.12	5054.734

**Figura 4.25:** Tabela comparativa da previsão e real 2017 do Gasóleo

Modelos	EQM	EAM	EPAM
ARIMA (0,1,0)	20.88707	890.3347	65.13942
ARIMA (1,1,0)	21.96484	880.9226	85.03498

**Figura 4.26:** Cálculo dos erros de previsão do gasóleo

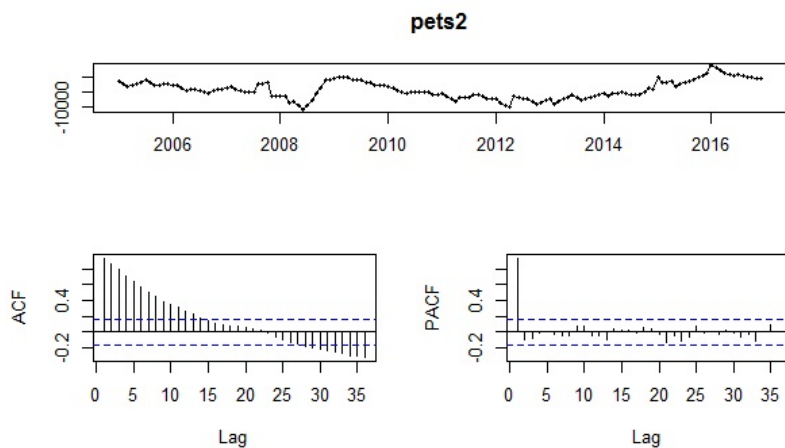
Note agora que depois da diferenciação, temos uma série estacionária. Ao analisarmos as FAC e FACP a questão seguinte é qual a estrutura do(s) modelo(s) do tipo ARMA e quais as suas ordens.

Depois do exposto, considerou-se dois possíveis modelos que podem ser adequados para descrever a série original e que são: ARIMA(0, 1, 0) e ARIMA(1, 1, 0).

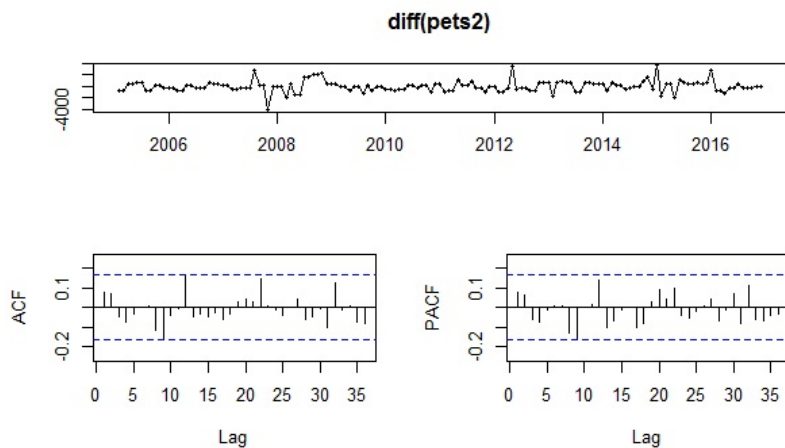
Após a escolha dos possíveis modelos foram, então, estimados os seus parâmetros.

De seguida procedeu-se à significância dos parâmetros estimados, através dos respectivos intervalos de confiança para os parâmetros do modelo.

Seguidamente, e após a estimação e avaliação da significância dos parâmetros dos modelos



**Figura 4.27:** FAC e FACP dos diferenciais de preço do petróleo até 2016



**Figura 4.28:** Série dos diferenciais de preço do petróleo transformada

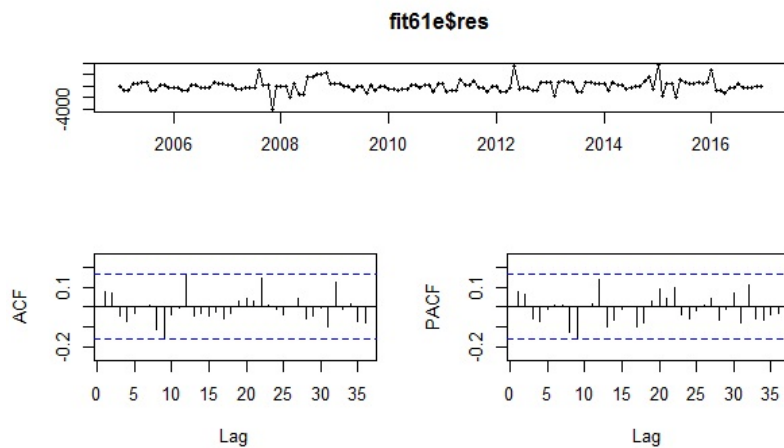
ARIMA(0, 1, 0) e ARIMA(1, 1, 0) , vai proceder-se à análise dos resíduos de modo a avaliar a adequação dos modelos.

Ao analisar os resíduos, da série e as respectivas FAC e FACP, tendo em conta as características que devem ter os resíduos (ver subsecção 3.5.3.E), os mesmos são aceitável. Neste sentido, não se rejeita a hipótese de os resíduos se comportarem como um ruído branco.

#### Testes aos resíduos do modelo ARIMA(0, 1, 0)

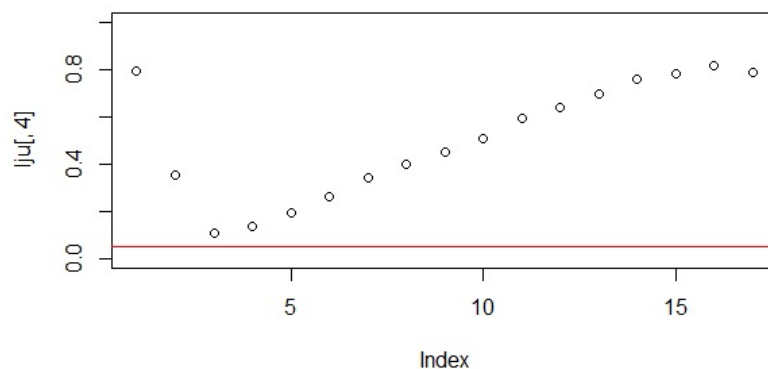
Para o nível de significância de 5%.

- Teste de Kolmogorov-Smirnov obteve um *p-value* de 0,037. Pelo que se rejeita a hipótese nula da normalidade dos dados.
- Teste t-student obteve um *p-value* de 0,932. Pelo que não se rejeita a hipótese nula da normalidade dos dados.
- Teste da diferença de sinais (Difference-Sign Test) obteve um *p-value* de 0,002. Pelo que se rejeita a hipótese nula da normalidade dos dados.



**Figura 4.29:** Resíduos do diferencial de preço do petróleo modelo ARIMA(0,1,0)

- De forma a testar a linearidade na média, realizou-se o teste de Ljung-Box, e o resultado é aceitável, visto que não existe nenhum elemento fora da banda de confiança, como se pode comprovar na figura 4.35 abaixo.
- Também realizou-se o teste de LiMcLeod (ver anexo 6.4) para identificar a não linearidade na média ou na variância, e não existe nenhum elemento fora da banda de confiança, sendo assim o resultado é aceitável.



**Figura 4.30:** Teste Ljung-Box realizado a série do petróleo modelo ARIMA(0,1,0) 2016

Também realizou-se o teste de LiMcLeod (ver anexo 6.4) para identificar a não linearidade na média ou na variância, e não existe nenhum elemento fora da banda de confiança, sendo assim o resultado é aceitável.

#### Testes aos resíduos do modelo ARIMA(1, 1, 0)

Também realizou-se os testes aos resíduos do modelo ARIMA(1, 1, 0) e obteve-se os seguintes resultados:

Para o nível de significância de 5%.

- Teste de Kolmogorov-Smirnov obteve um *p-value* de 0,038. Pelo que se rejeita a hipótese nula da normalidade dos dados.

- Teste t-student obteve um *p-value* de 0,938. Pelo que não se rejeita a hipótese nula da normalidade dos dados.
- Teste da diferença de sinais (Difference-Sign Test) obteve um *p-value* de 0,472. Pelo que não se rejeita a hipótese nula da normalidade dos dados.
- De forma a testa a linearidade na média, realizou-se o teste de Ljung-Box, e o resultado é aceitável, visto que não existe nenhum elemento fora da banda de confiança, como se pode ver no anexo 6.6.
- Também realizou-se o teste de LiMcLeod (ver anexo 6.5) para identificar a não linearidade na média ou na variância, e não existe nenhum elemento fora da banda de confiança, sendo assim o resultado é aceitável.

Ao comparar os resultados dos testes acima abordados, não fica bem explícito qual dos dois teve melhores resultados. Portanto, vai-se utilizar os critérios da subsecção 3.5.4 para identificar o melhor modelo.

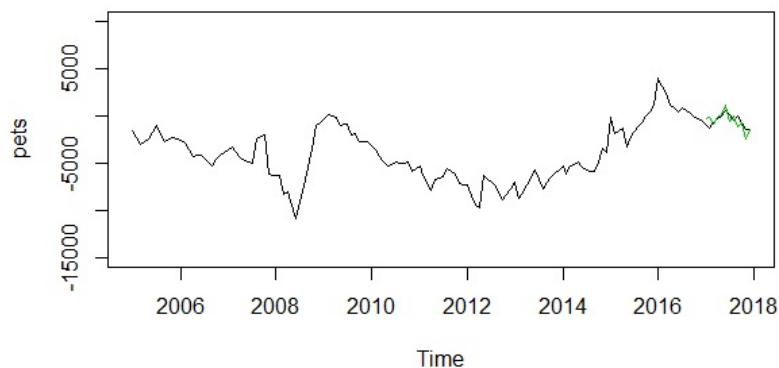
### Critérios de selecção de modelos

Modelo	AIC	AICc	BIC
ARIMA(0, 1, 0)	2378.48	2378.51	2381.45
ARIMA(1, 1, 0)	2379.61	2379.69	2385.53

**Tabela 4.3:** Critérios de comparação relativos ao ajustamento dos diferenciais de preço do petróleo

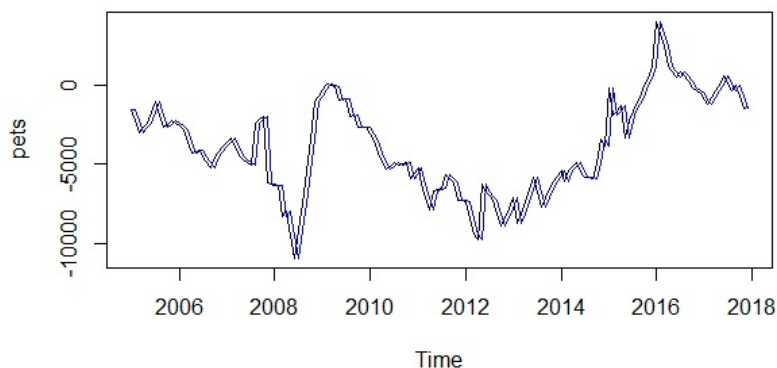
Após a comparação dos resultados dos critérios AIC, AICC e BIC dos dois modelos, e tendo em conta que ambos têm o mesmo número de parâmetros (princípio da parcimónia). Neste caso, o modelo ARIMA(0, 1, 0) é o que apresenta os menores valores e, portanto, é o modelo que se vai considerar como o melhor para descrever os dados.

Como já foi mencionado no início, de modo a avaliar a qualidade das previsões retirou-se o ano de 2017 (chamada série de teste) e ajustou-se um modelo à série dos diferenciais de preço do gasóleo entre os anos de 2005-2016. No início, foi retirado da série o ano de 2017, para depois compará-lo com a previsão deste mesmo ano, de forma a comprovar a capacidade preditiva do modelo, e pode-se ver através do gráfico abaixo, que a previsão do modelo é aceitável.



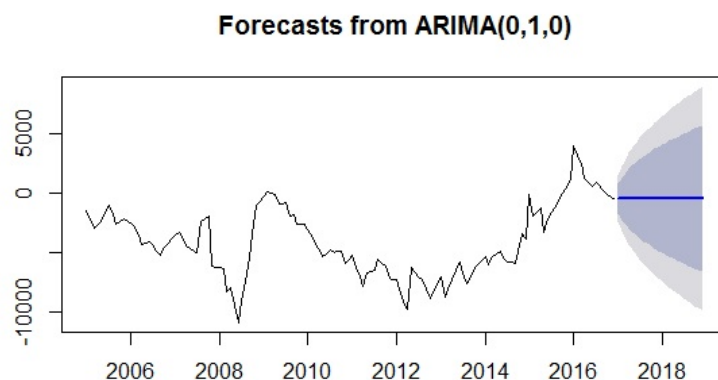
**Figura 4.31:** Gráfico comparativo do real e previsão 2017 de petróleo do modelo ARIMA(0, 1, 0)

De seguida apresentamos o gráfico do ajustamento modelo aos dados originais



**Figura 4.32:** Ajustamento do modelo  $ARIMA(0, 1, 0)$  aos diferenciais de preço da gasolina. Dados originais (a preto) e ajustados (a azul)

De acordo com o gráfico acima, pode-se ver que o modelo conseguiu descrever a série. Contudo, tal como foi referido no Capítulo 3, a qualidade das previsões também é um bom indicador sobre a qualidade do modelo ajustado, pelo que vamos agora fazer previsões através do modelo  $ARIMA(0, 1, 0)$  ajustado.



**Figura 4.33:** Previsão dos diferenciais de preço do petróleo para 2018-2019 modelo  $ARIMA(0, 1, 0)$

Através dos gráficos, pode ver que as previsões para os anos de 2018 e 2019 e as respectivas bandas de confiança a 80% e 95% (ver subsecção 3.6).

Foram também, calculada as medidas de EQM, EAM e EPAM para avaliar a performance preditiva dos modelos, ver subsecção 3.6.1.

## 4.5 Beneficiário da política de controlo do preço de combustível 2001-2017

Como um dos objectivos desta dissertação, também era averiguar o beneficiário de política que foi implementada a vários anos por sucessivos governos. Nesses cálculos, foram multiplicados por importação, as quantidades importadas pelos diferenciais de preço e, pode-se ver na figura seguinte.

Data	Real	Mean	Previsão
Jan 2017	-947.19	-511.43	1377.59647
Feb 2017	-1226.89	-511.43	1343.13252
Mar 2017	-778.93	-511.43	1644.35786
Apr 2017	-330.96	-511.43	1544.23912
May 2017	-16.72	-511.43	2096.72702
Jun 2017	616.25	-511.43	1127.17529
Jul 2017	115.53	-511.43	299.49471
Aug 2017	-385.19	-511.43	-418.45133
Sep 2017	-12.60	-511.43	-923.19374
Oct 2017	-737.92	-511.43	85.11771
Nov 2017	-1463.24	-511.43	-408.88651
Dec 2017	-1402.86	-511.43	-1869.74767

Figura 4.34: Tabela comparativa da previsão e real 2017 do petróleo

Ano	Gasolina			Gasóleo			Petróleo		
	Quant. Importada	Dif. Preço	Quant. X Preço	Quant. Importada	Dif. Preço	Quant. X Preço	Quant. Importada	Dif. Preço	Quant. X Preço
2001	5.981.196,00	9.676,00	6.528.160.472,00	16.357.390,00	7.899,00	16.601.783.143,00	4.562.229,00	3.055,00	2.503.336.294,00
2002	5.858.154,00	2.030,00	2.022.338.560,00	13.219.000,00	4.328,00	9.739.165.078,00	6.476.964,00	876,00	1.020.797.351,00
2003	5.271.832,00	4.019,00	4.266.443.166,00	14.852.981,00	22,00	-22.511.471,00	3.947.586,91	-1.474,00	-462.552.135,98
2004	6.910.055,00	-4.953,00	-4.713.116.022,00	35.424.874,00	-1.387,00	-15.161.953.056,00	15.677.586,00	-10.447,00	-29.863.133.109,00
2005	5.702.476,00	-12.586,00	-10.902.795.153,00	14.975.538,00	-1.673,00	-2.516.400.894,00	7.132.935,00	-12.861,00	-15.658.916.529,00
2006	6.002.365,00	-30.178,00	-25.388.668.427,00	17.519.276,00	-13.093,00	-33.450.905.417,00	6.037.055,00	-27.500,00	-23.750.819.745,00
2007	15.568.535,00	-23.598,00	-90.793.785.078,00	17.815.489,00	-10.324,00	-27.093.328.082,00	5.041.201,00	-28.045,00	-21.111.538.162,00
2008	8.284.356,00	-33.669,00	-39.147.863.401,00	16.555.889,00	-31.047,00	-74.430.938.200,00	9.055.774,00	-45.574,00	-57.142.808.681,00
2009	5.135.588,00	7.771,00	8.664.923.199,00	22.924.040,00	28.330,00	69.259.497.379,00	10.440.173,00	-11.333,00	-13.211.148.567,00
2010	8.272.378,00	-33.586,00	-39.173.223.450,00	20.656.217,00	-15.068,00	-44.877.674.964,00	9.820.732,00	-33.949,00	-47.729.440.880,00
2011	7.016.881,00	-46.192,15	-40.488.758.619,01	26.332.850,00	-30.685,31	-101.810.163.771,16	13.942.708,00	-51.675,68	-85.983.492.980,11
2012	6.320.560,00	-61.860,19	-58.894.193.677,35	20.508.606,00	-48.801,74	-137.124.101.605,97	8.237.742,00	-56.789,07	-67.741.722.596,37
2013	8.131.781,00	-50.846,96	-50.975.240.377,65	25.683.321,00	-45.905,37	-146.369.035.666,73	8.514.507,00	-57.290,75	-60.815.696.311,09
2014	9.462.883,00	-34.514,62	-30.554.752.534,78	37.486.737,00	-27.160,68	-98.166.701.640,23	9.763.531,00	-51.111,41	-50.654.978.765,95
2015	16.959.924,00	24.965,54	16.392.117.174,84	36.817.938,00	24.516,00	82.975.870.628,41	11.011.469,00	-9.754,55	-12.539.086.389,80
2016	17.567.492,00	52.125,03	91.782.793.982,89	34.845.754,00	47.456,01	182.035.990.206,40	20.575.175,00	9.699,18	14.764.225.862,87
2017	10.100.225,00	20.609,93	23.019.092.178,49	32.864.317,00	26.467,23	100.138.839.402,92	7.724.450,00	-5.169,40	-4.426.320.761,55
Total	148.546.681,00	-210.787,42	-238.356.528.006,57	404.840.217,00	-86.126,86	-220.272.568.930,36	157.961.817,91	-389.343,68	-472.803.296.105,98
Total Quant. X Preço			-931.432.393.042,91						
Total Quant. X Preço	euro	câmbio 24500 STD	38.017.648,69 Euro						

Figura 4.35: Beneficiários da política do diferencial de preço

Os valores estão expressos em dobras antigas, ou seja, foram retirados 3 dígitos, neste sentido: Total geral em antigas dobras = 931.432.393.042,91 STD. Total geral em novas dobras = 931.432.393,04 STD. Total geral em euro com o câmbio 24,5 do dia 29/11/2018 = 38.017.648,69 euro

Valores Positivos - O beneficiário é o Estado santomense  
Valores Negativos - O beneficiário é o consumidor final

Tendo em conta os resultados dos cálculos, o maior beneficiário desta política de controlo de preço dos combustíveis implementada entre 2001-2017, é o consumidor final, visto que comprou maior quantidade de produtos abaixo do preço real.

NOTA: Este valor (total geral) é aproximado, visto que faltam alguns factores que deveriam ser tomados em conta, que são os seguintes:

- O preço do gasóleo que se vende a EMAE (empresa de água e electricidade) diferem dos outros clientes;
- As quantidades do Jet-A1 não é afetada pelo diferencial do preço, mas está incluída no cálculo;
- A terceira questão, é que existem algumas facturas (principalmente dos últimos anos) do fornecedor (Sonangol, EP) que foram tomadas em consideração, mas são provisórias.

Não obstante as considerações acima mencionadas, os resultados desses cálculos são aceitáveis, visto que o objetivo é identificar o beneficiário e não a magnitude de afectação desta política.



# 5

## **Conclusões**

O caso de estudo em que a presente dissertação se insere descreve o problema do diferencial de preço que é causado pela política de controlo dos preços dos combustíveis adaptada pelo governo de São Tomé e Príncipe, antes do início desta dissertação. Esta política consistia em alterar menos possível os preços de venda ao público, independentemente das flutuações dos preços a nível internacional, mesmo estando estes sujeitos aos diferenciais de preços (positivos ou negativos) que esta política implica. O preço do combustível é uma variável que tem uma grande influência na economia de São Tomé e Príncipe, devido à dependência que muitos sectores têm destes mesmos produtos.

O objectivo da presente dissertação consistiu, em primeiro lugar, na implementação de modelos de Séries Temporais aos diferenciais dos preços da gasolina, do gasóleo e do petróleo e, posteriormente, à previsão (através dos melhores modelos ajustados) do diferencial de preço para cada combustível, para aos anos futuros.

Em segundo lugar, através das análises das quantidades importadas e do diferencial de preços, saber qual o beneficiário desta política de controlo de preço praticada pelo Estado santomense e propôr medidas de melhoria para ambas as partes, estado santomense e empresa ENCO, SARL.

Para tal, foram recolhidos dados do Mapa de Preços disponibilizados pela ENCO, SARL, de forma obter os dados da importação e formação de preços, no período de 17 anos, entre 2001 e 2017. De seguida foi necessário proceder-se ao tratamento destes dados já que os mesmos não eram igualmente espaçados. A forma que nos pareceu mais adequada, de modo a não perder informação e a não alterar o padrão original dos dados foi transformar estes em dados mensais.

Após a adequação dos dados para a aplicação dos modelos de Séries Temporais, nomeadamente dos modelos ARIMA/SARIMA, procedeu-se, então à análise descritiva dos dados e, posteriormente ao ajustamento dos modelos às três séries de dados. E, por último, com base nos modelos ajustados fez-se a previsão dos diferenciais dos preços para os anos de 2018-2019.

Através da análise descritiva dos gráficos foi possível observar algumas características nas séries. Gasolina: No período de 2001 a princípio de 2005 os valores do preços real (PR) e Preço de venda (PV) eram muito próximo, fazendo com que o diferencial de preço tivesse valores muito pequenos, isso reflete a preocupação do governo/s deste período, em minimizar os efeitos do diferencial de preço, visto que os mesmo/s actualizavam constantemente os PV.

Após este período começa-se a notar o abrandamento da política de minimizar o diferencial de preço, visto que as diferenças entre o PR e PV começam a ser cada vez maiores, e a sua tendência crescente mais acentuada. Neste período, mais concretamente no ano de 2008, houve um incremento acentuado dos preços dos combustíveis originado pela crise financeira internacional.

Nesta ultima fase, começa-se em princípio de 2011 com um aumento do preços de combustíveis. onde a gasolina sofreu um aumento de 18,18%, gasóleo 19,44% e petróleo em 33,33%. a política neste fase não se preocupou com o diferencial do preço, visto que os PR flutuava e PV continuava constante.

Também na análise descritiva das séries, apurou-se que todas tinham componente tendencial, ou seja, uma tendência crescente, mas não se encontrou componente sazonal em nenhuma das série. Esta tendência está relacionada com as flutuações do preço de combustíveis a nível internacional que não estava sendo bem acompanhada pelas políticas aplicada pelos sucessivos governos.

Os modelos estatísticos utilizados nesta dissertação são modelos ARIMA/SARIMA, com os mesmos foi possível descrever (ajustar) e também prever valores futuros.

O principal objectivo da dissertação, é encontrar os melhores modelos que possam descrever e fazer previsão das séries para o ano 2018 e 2019. Os modelos estatísticos utilizados nesta dissertação são modelos ARIMA/SARIMA, com os mesmos foi possível descrever (ajustar) e também prever valores futuros das séries.

**Série da gasolina** Após analisar a série da gasolina, considerou-se os modelos SARIMA(0, 1, 0) × (0, 0, 1)<sub>12</sub> e SARIMA(0, 1, 0) × (1, 0, 0)<sub>12</sub>, como os possíveis modelos mais adequados para realizar a previsão da série. Depois de realizar vários testes aos parâmetros de ambos os modelos, o modelo SARIMA(0, 1, 0) × (0, 0, 1)<sub>12</sub> foi o que apresentou melhores resultados, sendo assim, foi escolhido como o mais adequado para descrever e realizar a previsão da série da gasolina.

**Série do gasóleo** Da mesma forma que a série da gasolina, também considerou-se dois possíveis modelos para a série do gasóleo que são: ARIMA(0, 1, 0) e ARIMA(1, 1, 0). Após a realização de vários testes aos parâmetros dos modelos, não obstante os resultados dos modelos serem muito semelhantes, considerou-se o modelo ARIMA(0, 1, 0) como o melhor modelo para descrever e fazer a previsão da série.

**Série do petróleo** Al igual que a séries da gasolina e do gasóleo, considerou-se também dois

possíveis modelos para realizar a previsão da série do petróleo. Dada a semelhança das séries do gasóleo e do petróleo, os modelos considerados como adequados são os mesmos, ou seja, os modelos  $ARIMA(0, 1, 0)$  e  $ARIMA(1, 1, 0)$ . Após a realização de vários testes aos parâmetros dos modelos e, tendo em conta os resultados, considerou-se o modelo  $ARIMA(0, 1, 0)$  com o melhor para descrever a série, enquanto que o modelo  $ARIMA(1, 1, 0)$  é o mais adequado para fazer a previsão da mesma.

Um dos objetivos desta dissertação, também era averiguar entre o Estado santomense e o consumidor final, o beneficiário desta política de controlo de preço praticada durante o período de 2001-2017. Após os cálculos realizados aos dados de importação e do mapa de preço facultados pela ENCO, SARL foi possível concluir que o maior beneficiário e o consumidor final, visto que compraram maior quantidade de produtos abaixo do PR.

A política mais adequada, seria actualizar os preços de venda ao público sempre que os preços de compra sofressem grandes flutuações a nível internacional, de forma a manter o diferencial de preço o mais baixo possível.

É relevantes mencionar as principais melhorias que podem ser agregadas em estudos futuros dentro do mesmo tema desta dissertação. Essas melhorias podem ser: incluir no estudo a variável - câmbio do dólar americano - como uma variável que pode influenciar os preços; incluir também a opinião dos peritos na previsão dos dados; tomar em consideração nos cálculos todos os factores que influenciam na quantidade e no diferencial de preço dos produtos, para se saber exatamente a magnitude das políticas de controlo de preço praticada em São Tomé e Príncipe.

# 6

## Anexos

---

<b>6.1 Anexos</b> .....	<b>77</b>
-------------------------	-----------

## 6.1 Anexos

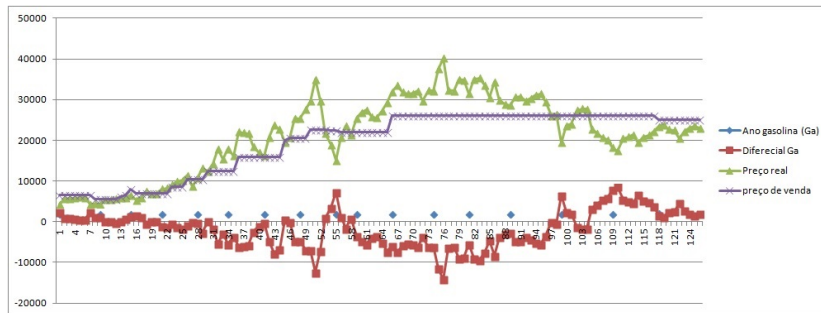


Figura 6.1: Gráfico da série original da gasolina (Diferencial de preço, preços real e de venda)

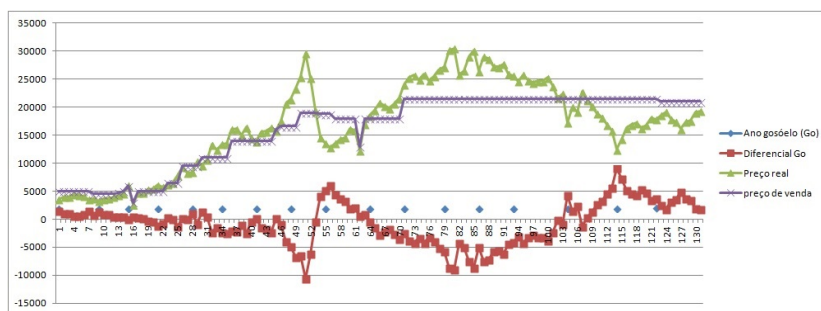


Figura 6.2: Gráfico da série original do gasóleo (Diferencial de preço, preços real e de venda)

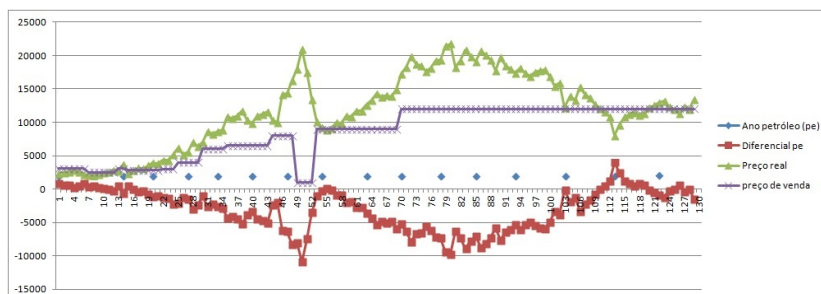
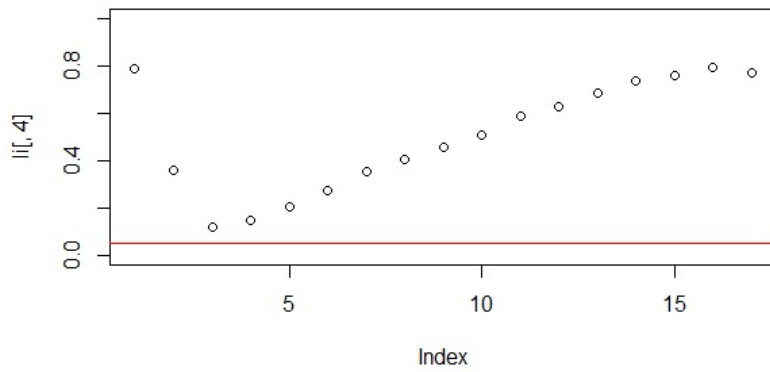
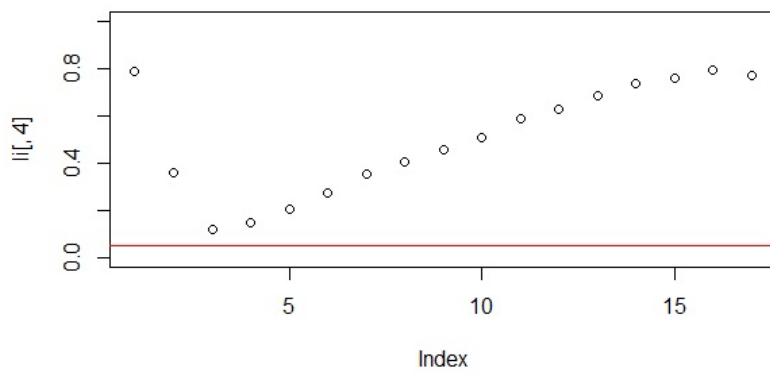


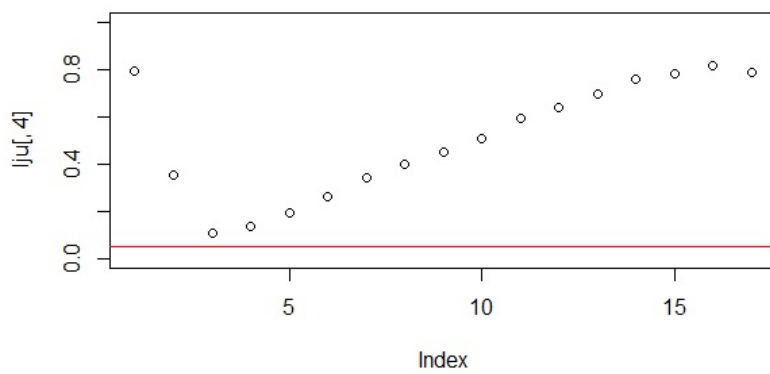
Figura 6.3: Gráfico da série original do petróleo (Diferencial de preço, preços real e de venda)



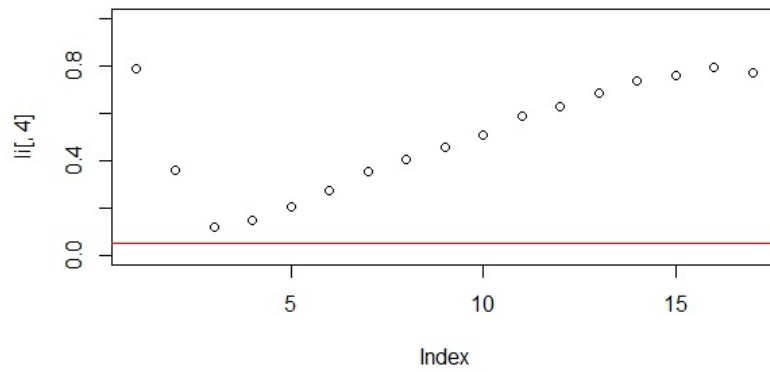
**Figura 6.4:** Teste Li-McLeod para a série do petróleo modelo ARIMA(0, 1, 0)



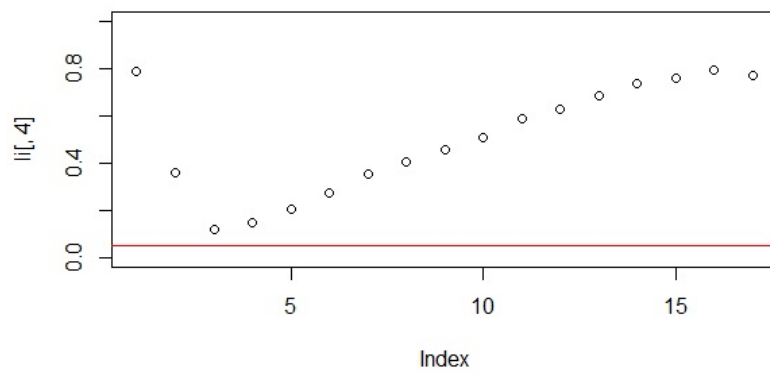
**Figura 6.5:** Teste Li-McLeod para a série do petróleo modelo ARIMA(1, 1, 0)



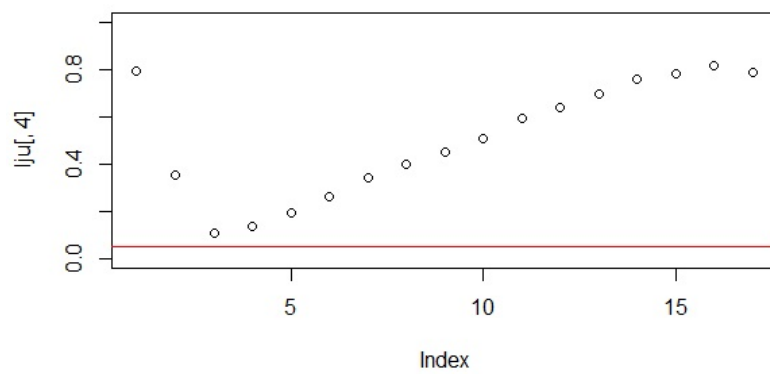
**Figura 6.6:** Teste Ljung-Box para a série do petróleo modelo ARIMA(1, 1, 0)



**Figura 6.7:** Teste Li-McLeod para a série do gasóleo modelo(0, 1, 0)



**Figura 6.8:** Teste Li-McLeod para a série do gasóleo modelo(1, 1, 0)



**Figura 6.9:** Teste Ljung-Box para a série do gasóleo modelo(1, 1, 0)

# Referências bibliográficas

- [1] Agência Nacional de Petróleo de São Tomé e Príncipe. <https://www.anp-stp.gov.st> (Consultado a 15 fevereiro de 2017).
- [2] Alvarenga, S., Vieira, K. e Fialho, P. (2017). *Demanda por gasolina: um estudo de caso para uma rede de postos de combustíveis*. Revista do CEPE. Santa Cruz do Sul, n. 46, pp. 149-165. ISSN on-line: 1982-6729. DOI: 10.17058/cepe.v0i46.9063. <https://online.unisc.br/seer/index.php/cepe/article/download/9063/6870>.
- [3] Africahrpt (20 Junho 2017). *Petróleo e Gás: A África em busca de novos projetos*. Africa Hr Solutions (Blog) <http://africa-hr.com/pt/2017/06/20/petroleo-e-gas-a-africa-em-busca-de-novos-projetos> (consultado em 05 de setembro de 2018).
- [4] Balogun, S. e Ogunleye, M. (2015). *On the time series modelling of crude oil exportation in Nigeria*. European Journal of Academic Essays 2(3), pp. 1-12. ISSN (online): 2183-1904. ISSN (print): 2183-3818 <http://euroessays.org/wp-content/uploads/2015/04/EJAE-346.pdf>.
- [5] Banco Central de São Tomé e Príncipe. [www.bcstp.st](http://www.bcstp.st) (Consultado a 10 fevereiro de 2018).
- [6] Brockwell, P. e Davis, R. (1991). *Time series: theory and method*. Second Edition, Springer-Verlag, New York.
- [7] Chaize, T. (2010). *China e Africano de petróleo*. Em <http://www.dani2989.com/matiere1/chinaafrica0810pt.htm> (Consultado a 05 de Setembro de 2018).
- [8] Cowpertwait, P. e Metcalfe, A. (2009). *Introductory time series with R*. Springer.
- [9] Cristina, V. (2001). *Aprendendo Latex*. Labmac-UFPA. [http://www.lac.inpe.br/~margarete/download/MET200-0/apostila\\_latexpdf.pdf](http://www.lac.inpe.br/~margarete/download/MET200-0/apostila_latexpdf.pdf).
- [10] Cryer, J. e Chan, Kung-Sik (2008). *Time series analysis: with applications in R*. Springer.
- [11] Ehlers, R. S. (2007). *Análise de Séries Temporais*. Quarta edição, Universidade Federal do Paraná, <http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf>.
- [12] Esteves, P. S. e Neves, P. D. (2004). *Efeitos económicos das flutuação do preço do petróleo*. Boletim Económico, Banco de Portugal.
- [13] Gary, I. e Karl, T. (2003). *O fundo do barril: O 'boom' do petróleo em África e os pobres*. Catholic Relief Services. [https://www.laohamutuk.org/OilWeb/Bground/Africa/CRS%20Africa%20report/oil\\_report\\_full\\_port.pdf](https://www.laohamutuk.org/OilWeb/Bground/Africa/CRS%20Africa%20report/oil_report_full_port.pdf)
- [14] Gomes, D. (2005). *Séries temporais. Sebenta da cadeira de séries temporais*. Universidade de Évora. Portugal.
- [15] International Energy Agency. [www.iea.org](http://www.iea.org) (Consultado a 08 outubro de 2018).
- [16] International Energy Forum (2018). *Comparative analysis of monthly reports on the oil market*. An International Energy Forum Publication. Riyadh, Saudi Arabia. [https://www.ief.org/\\_resources/files/comparative-analysis/august-2018-comparative-analysis.pdf](https://www.ief.org/_resources/files/comparative-analysis/august-2018-comparative-analysis.pdf)



- [17] Instituto Nacional de Estatística de São Tomé e Príncipe. <https://www.ine.st/> (Consultado a 12 fevereiro de 2018).
- [18] Kilian, L. e Park, C. (2009). *The impact of oil price shocks on the U.S. stock market*. International Economic Review, Wiley Online Library, v.50, n.4, pp. 1267-1287. <https://doi.org/10.1111/j.1468-2354.2009.00568.x>.
- [19] Makridakis, S. e Hibon, M. (1995). *ARMA models and the Box Jenkin methodology*. working paper in the INSEAD Working Paper Series. <https://autobox.com/makridakis.pdf>.
- [20] Mártil, I. (2016). *La historia del precio del petróleo: Subidos a una montaña rusa*, *EconoNuestra*, <https://blogs.publico.es/econonuestra/2016/02/13/la-historia-del-precio-del-petroleo-subidos-a-una-montana-rusa/> (consultado em 14 de agosto de 2018).
- [21] Moura, B. (2014). *O papel do açúcar e da gasolina na determinação dos preços de etanol no Brasil* (Tese de Mestrado). Universidade Federal do Rio de Janeiro.
- [22] Murteira, B., Muller, D. e Turkman, F. (1993). *Análise de sucessões cronológicas*. McGraw-Hill, Portugal.
- [23] National Bureau of Economic Research (2008). *Understanding crude oil prices*. Cambridge, EUA. <https://www.nber.org/digest/mar09/w14492.html> (consultado em 05 de setembro de 2018).
- [24] Organization of the Petroleum Exporting Countries. <https://www.opec.org> (Consultado a 15 fevereiro de 201).
- [25] Pacheco, P. (2001). *Notas de séries temporais*, Instituto Superior Técnico, Lisboa. <https://fenix.tecnico.ulisboa.pt/downloadFile/3779571248833/Notas.pdf>.
- [26] Relatório do Fundo Monetário Internacional n.14/2 (17 Janeiro de 2014). <https://www.imf.org/external/lang/Portuguese/pubs/ft/scr/2014/cr1402p.pdf>.
- [27] Shumway, R. e Stoffer, D. (2006). *Time series analysis and its applications: with R examples*. Springer.
- [28] Tularam, G. e Saeed, T. (2016). *Oil-Price Forecasting Based on Various Univariate Time-Series Models*. King Abdulaziz University. <https://www.scirp.org/journal/PaperInformation.aspx?paperID=66527>.
- [29] Wei, W. (2006). *Time series analysis: univariate and multivariate methods*. Second Edition, Library of Congress Cataloging- in-Publication Data, EUA. ISBN 0-321-32216-9. [https://www.researchgate.net/publication/236651810\\_Time\\_Series\\_Analysis\\_Univariate\\_and\\_Multivariate\\_Methods\\_2nd\\_edition\\_2006](https://www.researchgate.net/publication/236651810_Time_Series_Analysis_Univariate_and_Multivariate_Methods_2nd_edition_2006)
- [30] Ye, M., Zyren, J. e Shore, J. (2002). *Forecasting crude oil spot price using OECD petroleum inventory levels*. International Advances in Economic Research, Springer, v.8, n.4, pp. 324-333.
- [31] Zoubir, L. (2017). *A brief history of time series analysis*. Department of Statistics, Stockholm University. <https://www.statistics.su.se/english/research/time-series-analysis/a-brief-history-of-time-series-analysis-1.259451>.